

**Appraisal Theory as a Linguistic Tool for the Analysis of Market
Research Interview Data**

by

Russell Wilson

Submitted in accordance with the requirements for the degree of
PhD

The University of Leeds
School of Modern Languages & Cultures

May, 2011

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Acknowledgements

I would first like to place on record my thanks to my supervisor, Judy Delin, for providing so much support and assistance throughout this process. Our paths first crossed in 2001 and she has provided guidance and support both before and during my PhD studies. I would also like to give thanks to Alice Deignan for taking over all of these duties for the latter half of my studies and showing much enthusiasm for ensuring that I made it over the finishing line. Her generosity, both in terms of time and energy, meant I was never allowed to succumb to the idea that I might not finish this thesis. Sally Johnson and Lynne Cameron, who both also acted as supervisors at various points along the way, also deserve my thanks.

Much of the funding for this research was provided by the Faraday Group and I would also like to thank them for making this PhD a viable proposition. I am particularly indebted to Sue Hepworth and Stephen Worrall at Unilever for providing help and advice above and beyond the six months we spent working together. Without the facilities and advice that they and the rest of the CTI team at Seacroft provided, this research would not have been possible. I am delighted that in addition to the data, I also collected a wonderful mentor, in Sue Hepworth, who I continue to turn to for advice and guidance.

Many thanks also to Steve Lillford and Martin Thomas who provided many hours of interesting conversation and acted as wonderful sounding boards to a huge number of questions, thoughts and ideas. I would also like to thank my family for providing incredible support over the last four years with only very infrequent queries as to when I would get a 'proper' job.

Final thanks are reserved for my wonderful Emy. Despite spending so much of my life studying them, it is simply not possible to find words that adequately sum up the role that she has played in helping me to complete this thesis. Without her steady stream of encouragement, advice and support this thesis would not have been completed.

Abstract

The use of linguistics within market research is for the most part, marked by its absence. This is perhaps surprising given the potential it offers for analysing what people have said and what they might mean. Though the study of 'evaluation' has been approached from many different linguistic perspectives, previous work in this field has tended to focus on individual markers, rather than aiming to provide a fuller, more comprehensive account.

This thesis proposes that it is possible to combine approaches from Discourse and Conversation Analysis, with developments in the field of Systemic Functional Grammar, to gain a more inclusive understanding of the social and interactional influences that can determine how an evaluation is both formulated and articulated.

The data for this study was collected from thirty paired depth interviews, in the field of New Product Development. This data was transcribed and tagged using O'Donnell's (2007) CorpusTool software. It was then analysed using Martin and White's (2005) framework of Appraisal Theory, in conjunction with a scale developed from Brown and Levinson's (1987) Politeness Theory and Sinclair and Coulthard's (1975, 1992) work on teacher/ pupil interactions.

As a result of the analysis carried out in this study, two potential extensions to the Appraisal Theory Framework are suggested. These extensions are with regards to the relevance of the subject matter to the speaker making the evaluation, and the notion of neutral evaluations. In addition to taking an existing framework and developing it for a new purpose, this thesis also contributes to the wider understanding of 'evaluation', through the development of a Scale of Importance for individual turns, with regards to the 'weight' that should be assigned to them due to their place in the turn taking structure.

Table of Contents

ACKNOWLEDGEMENTS
ABSTRACT	1
TABLE OF CONTENTS	2
INDEX OF TABLES.....	6
INDEX OF FIGURES	7
1. INTRODUCTION AND REVIEW OF MARKET RESEARCH LITERATURE....	8
1.1 Communicators as Consumers.....	11
1.1.1 A Wanting Animal.....	12
1.1.2 The Consumer Decision Making Process	15
1.1.3 Products as Attributes	18
1.1.4 The Attitude-Behaviour Problem.....	22
1.1.5 The Theory of Planned Behaviour.....	23
1.1.6 Manipulating Consumer Feelings	28
1.2 Market Research	29
1.2.1 Market Research and Qualitative Interviews in the Development Process	30
1.2.2 Approaches to Analysing Qualitative Data.....	33
2. INTERPRETING EVALUATIVE LANGUAGE	36
2.1 Discourse	37
2.1.1 Conversation Analysis.....	38
2.1.1.1 Meaning in Conversation as a Joint Production.....	40
2.1.1.2 Turn Taking	41
2.1.1.3 Preference and Power.....	49
2.1.1.3.1 Preference.....	49
2.1.1.3.2 Power.....	52
2.1.1.4 Content and Non Content turns.....	58
2.1.1.5 Tag Questions	61
2.1.2 Politeness Theory.....	62
2.1.2.1 Face.....	64
2.1.3 Proposing a Scale of Inherent Strength Based on Structural and Social Factors	65
2.2 Grammar.....	68
2.2.1 Stance.....	71
2.2.2 Modality.....	75
2.2.3 Hedging.....	78
2.2.4 Evidentiality.....	80
2.3 Lexis.....	82
2.3.1 Semantic Prosody.....	84

2.4 Phonology and Non Verbal Communication.....	88
2.4.1 Prosody: Pitch, Tone and Pace.....	88
2.4.2 Non Verbal Communication.....	89
3. THEORETICAL FRAMEWORKS FOR THE ANALYSIS OF EVALUATION .91	
3.1 Systemic Functional Grammar	91
3.2 Appraisal Theory.....	93
3.2.1 Attitude.....	94
3.2.1.1 Affect	95
3.2.1.2 Judgement	97
3.2.1.3 Appreciation.....	99
3.2.2 Engagement	103
3.2.3 Graduation	107
3.2.4 Summary of Appraisal Theory.....	110
3.3 Summary of the Two Frameworks	111
4. PRELIMINARY METHODOLOGY AND PILOT STUDY	112
4.1 Issues in Interviews and Questionnaire Design.....	112
4.1.1 Interview Types	112
4.1.1.1 Standardised Interviews	113
4.1.1.2 Exploratory Interviews.....	115
4.1.2 Self-Report Methods.....	117
4.1.2.1 Triadic sorting (Kelly Grids).....	118
4.1.2.2 Free Sorting and Ranking.....	120
4.1.2.3 Direct Elicitation	120
4.2 Pilot Study	121
4.2.1 Participants of the Pilot Study	121
4.2.2 Methodology of the Pilot Study.....	123
4.2.3 Transcribing the Data from the Pilot Study	125
4.3 Analysis and Discussion	126
4.3.1 Attitude Discussed	127
4.3.2 Engagement Discussed	130
4.3.3 Graduation Discussed	132
4.3.4 Inscribed and Evoked Appraisal	134
4.3.5 Triadic Sorting Analysed.....	136
4.3.5.1 Points of Comparison.....	137
4.3.5.2 Problems with Triadic Sorting	138
4.3.5.3 Prototype Theory	140
4.4 Summary of the Pilot Study.....	142
5. FINAL METHODOLOGY AND MAIN STUDY	144
5.1 Introducing the Main Study	145
5.1.1 The Participants	145
5.1.1.1 The Unilever Panel.....	146
5.1.1.2 Skilled Participants	147

5.2 Methodology of the Main Study: A Four Stage Process	147
5.2.1 Stage 1: Initial Focus Groups.....	149
5.2.2 Stage 2: The CAD Sessions	151
5.2.3 Stage 3: Reassessing the Prototype.....	153
5.2.4 Stage 4: Paired Depth-Interviews	154
5.2.4.1 Triadic Sorting.....	156
5.2.4.2 Ranking.....	156
5.3 Transcribing the Data from the Main Study.....	157
5.4 Tagging the Data from the Main Study.....	157
5.4.1. The Appraisal Layer	162
5.4.1.1 Tagging Attitude	163
5.4.1.1.1 Tagging Affect	163
5.4.1.1.2 Tagging Judgement	164
5.4.1.1.3 Tagging Appreciation.....	166
5.4.1.1.4 Tagging Engagement.....	171
5.4.1.1.5 Tagging Graduation.....	173
5.4.2 The Turn-taking Layer.....	175
5.4.3 The Participant Layer.....	179
5.5 Inter-rater Reliability.....	179
5.5.1 Testing Reliability	181
5.5.2 Comparing the Tagging	181
6. RESULTS	183
6.1 Distribution of Positive and Negative Evaluations across Appraisal Theory Framework.....	185
6.1.1 Positive Evaluations for Egypt.....	186
6.1.2 Positive Evaluations for Mouse	188
6.1.3 Negative Evaluations for Egypt.....	190
6.1.4 Negative Evaluations for Mouse.....	192
6.2 Weakening and Strengthening of Evaluations in Relation to their Polarity	194
6.2.1 Weakening of Positive Evaluations about Egypt.....	195
6.2.2 Strengthening of Positive Evaluations about Egypt.....	196
6.2.3 Weakening of Positive Evaluations about Mouse.....	198
6.2.4 Strengthening of Positive Evaluations about Mouse	201
6.2.5 Weakening of Negative Evaluations about Egypt	202
6.2.6 Strengthening of Negative evaluations about Egypt.....	204
6.2.7 Weakening of Negative Evaluations about Mouse	206
6.2.8 Strengthening of Negative Evaluation about Mouse	207
6.3 Turn types in relation to weakening and strengthening strategies.....	209
6.3.1 Initiating and Dispreferred turns in relation to Positive Evaluations about Egypt.....	209
6.3.2 Initiating and Dispreferred turns in relation to Negative Evaluations about Egypt	213
6.3.3 Initiating and Dispreferred turns in relation to Positive Evaluations about Mouse.....	217
6.3.4 Initiating and Dispreferred turns in relation to Negative Evaluations about Mouse	220
7. DISCUSSION.....	223
7.1 Polarity In Relation to Ranking and Category	224
7.2 Weakening and Strengthening Strategies in Relation to Polarity	231

7.2.1 Weakening of Positive Evaluations	231
7.2.2 Strengthening of Positive Evaluations	232
7.2.3 Weakening of Negative Evaluations.....	240
7.2.4 Strengthening of Negative Evaluations.....	243
7.3 Weakening and Strengthening Strategies in Relation to Ranking	248
7.4 Shortcomings and Suggested Modifications.....	249
7.4.1 Engagement: Relevance.....	250
7.4.2 Neutral Evaluations	254
8. CONCLUSION.....	258
8.1 Applying linguistic approaches to commercial practice.....	261
8.2 Strengths and Limitation of the study	265
8.2.1 Contributions to the Knowledge	266
8.2.2 Limitations.....	266
8.3 Areas for Further Research.....	267
REFERENCES	269
APPENDICES.....	298

Index of Tables

Table 1 Claimed features of Qualitative and Quantitative methods (Halfpenny 1979: 799)	32
Table 2 Six different affect groups (Biber and Finegan 1989)	74
Table 3 The different ways in which Affect can be realised (Martin 2003)	95
Table 4 Examples of Affect across the three main categories of emotion (Martin 2003)	96
Table 5 System of Judgement (Martin and White 2003: 53)	98
Table 6 A table to show the split between Positive and Negative Force	133
Table 7 The split between Positive and Negative FOCUS	134
Table 8 Distribution of Positive Evaluations about Egypt Across the Appraisal Theory Framework	186
Table 9 Distribution of Positive Evaluations about Mouse Across the Appraisal Theory Framework	188
Table 10 Distribution of Negative Evaluations about Egypt Across the Appraisal Theory Framework	190
Table 11 Distribution of Negative Evaluations about Mouse Across the Appraisal Theory Framework	192
Table 12 Distribution of different weakening strategies for positive evaluations about Egypt	195
Table 13 Distribution of different strengthening strategies for positive evaluations about Egypt	197
Table 14 Distribution of different weakening strategies for positive evaluations about Mouse	198
Table 15 Distribution of different strengthening strategies for positive evaluations about Mouse	201
Table 16 Distribution of different weakening strategies for negative evaluations about Egypt	203
Table 17 Distribution of different strengthening strategies for negative evaluations about Egypt	204
Table 18 Distribution of different weakening strategies for negative evaluations about Mouse	206
Table 19 Distribution of different strengthening strategies for negative evaluations about Mouse	207
Table 20 Distribution of Turn Types for Positive Evaluations about Egypt	210
Table 21 Distribution of Turn Types for Negative Evaluations about Egypt	214
Table 22 Distribution of Turn Types for Positive Evaluations about Mouse	217
Table 23 Distribution of Turn Types for Negative Evaluations about Mouse	220
Table 24 Distribution of Positive and Negative Evaluations about Egypt across the Appraisal Theory Framework	227
Table 25 Distribution of Positive and Negative Evaluations about Mouse across the Appraisal Theory Framework	228
Table 26 Frequency of Weakening and Strengthening Markers across both Stimuli	247

Index of Figures

Figure 1 Maslow's hierarchy of needs (Maslow 1970)	13
Figure 2 A hierarchy of consumer needs (Jordan 2000: 6)	13
Figure 3 The Consumer Decision Making Process (Blackwell et al. 2005)	16
Figure 4 The standard symbol to show that a product is made from recycled materials	20
Figure 5 Theory of Planned Behaviour (Ajzen 1991)	24
Figure 6 A diagrammatic representation of Sinclair and Coulthard's Initiation-Response (Farooq 1999: 31)	44
Figure 7 Basic Speech Roles	92
Figure 8 The affectual basis of judgement and appreciation (White 1998: 154)	103
Figure 9 The seven different bottle shapes used in the experiment	124
Figure 10 Four Stages of Development	148
Figure 11-Sketch 1 Figure 12 Sketch 2 Figure 13 Sketch 3	150
Figure 14 Sketch 4 Figure 15 Sketch 5	150
Figure 16 Design A Figure 17 Design B Figure 18 Design C	151
Figure 19 Design D Figure 20 Design E	151
Figure 21 Still From CAD Session	152
Figure 22 Screen shot of the Appraisal Network on CorpusTool software	160
Figure 23 Screen shot of the Turn-type Network on CorpusTool software	161
Figure 24 Screen shot of the Participant Network on CorpusTool software	161
Figure 25 Stimulus Mouse Figure 26 Stimulus Egypt	183
Figure 27 The Attitude: Polarity branch of the Appraisal Framework (Martin and White 2005).	255
Figure 28 Amended Attitude: Polarity branch of the Appraisal Framework.	256

1. Introduction and review of market research literature

The aim of this research is to develop a deeper understanding of linguistic evaluations. In the introduction to their influential book, 'Evaluation in text: Authorial Stance and the Construction of Discourse' Hunston and Thompson (2003:5) define evaluation as:

'the broad cover term for the expression of the speaker or writer's attitude or stance towards, viewpoint on, or feelings about the entities or propositions that he or she is talking about.'

This wide definition alludes to the complexity of the different factors that have an impact on evaluations and suggests that a variety of linguistics approaches might be employed to help understand them. The way in which different linguistic resources have been used in relation to understanding evaluation is detailed below.

Much of the funding for this study was provided by the Faraday Group. This is an organization comprised of Unilever, Procter & Gamble, Nestle, Diageo, Masterfoods and Britvic. Between them, they are responsible for many leading consumer brands and products such as Lynx, Gillette, Dove, Quality Street, Guinness, Mars and Pepsi. As part of this industry sponsored, Case Studentship, I spent six months on placement with Unilever's Consumer Insight Technical team, working on a project to explore new packaging design solutions. As a result, language pertaining to the areas of shape, packaging design and new product development are of particular interest. The aim of this placement was to provide the setting for the data collection process to take place, within the context of an active market research team. The data collection process entailed the design and implementation of thirty hours worth of research interviews, with consumers from the Unilever Panel. The stimuli for these interviews were prototype designs for a deodorant pack that were also developed as part of the research methodology. This whole process is discussed in detail in Chapter 5. Images of the two packs that are discussed in detail are shown in the introduction to Chapter 6.

One of the central arguments, and potential applications of this thesis, is that using analytic tools from the field of linguistics will allow the researcher to view their data from a different perspective, which may prove useful in gaining different insights into respondents' opinions. Whereas previously it was considered enough to have gathered information about consumers, it is becoming apparent that what is of real importance is what that information means. As Smith and Fletcher (2004: 2) argue 'Information was once power. But today, the power lies in interpreting what information *really* means'. In addition to broadening understanding of how evaluations are constructed and presented, findings from this research could potentially be used to aid the design of new research methodologies. This might then facilitate the creation of new packaging designs that better meet the needs of the consumer. A greater understanding of the social and interactional influences effecting evaluative talk could also prove useful in training research moderators and managing client/ researcher relations. This thesis aims to answer the following questions:

- 1) How can linguistic approaches help understand inherent levels of strength or weakness in evaluative talk?
 - a) Is there a link between the distribution of strengthening or weakening markers and the explicit ratings given in the ranking exercise?
 - b) Is there a link between the polarity of an evaluation and the distribution of strengthening or weakening markers?
 - c) Is there a link between the overall ranking of stimulus and the distribution of positive and negative evaluations?

- 2) What are the social and interactional considerations that need to be taken into account when attempting to understand linguistic evaluations?
 - a) Are different categories of the Appraisal Theory Framework more likely to attract positive or negative evaluations and does this have

consequences with regards to the participants' overall evaluation of the pack?

- 3) How could manufacturers of products use knowledge of these considerations when designing products and packaging?

Chapter 1 provides an overview of issues relating to consumer behaviour, new product development and consumer insight research, providing the background against which this research is carried out. I will then review, in Chapter 2, the current literature on evaluative language across four different levels of language:

- Discourse
- Grammar
- Lexis
- Phonology

Having introduced and explored different approaches to understanding evaluation, Chapter 3 assesses two frameworks of analysis that attempt to set out clear methodologies for analysing evaluative language. Firstly, Systemic Functional Grammar (SFG) is introduced and explored as an approach to analysing data elicited from market research interviews. Appraisal Theory, which developed out of SFG, is then introduced and discussed in greater detail with regards to its relevance and suitability for helping to answer the questions set out above. Chapter 4 discusses methodological issues and introduces a small pilot study before the main study is introduced and discussed in Chapter 5. The results of the analysis of the data from the main study are presented in Chapter 6 which also includes some preliminary discussion. The main discussion is presented in Chapter 7 before a conclusion and exploration of further areas of study is carried out in Chapter 8.

1.1 Communicators as Consumers

The aim of this research is to better understand the ways in which evaluative language is used, the main purpose of this being to contribute towards the development or improvement of research methodologies within the field of consumer insight and product design. Existing approaches to this research process are detailed below in section 1.2. Any developments might then be used to aid the design and manufacture of products which better meet the demands of the people who buy them.

For brand owners, the ability to gain insights into how their consumers relate to and interact with their product packaging is important because whenever a consumer comes into contact with a business, through their products; they are likely to form an impression of that company (Carlzon 1989). Any impressions formed are, it is suggested, likely to affect the subsequent likelihood of the consumer buying the product or using the service. Carlzon's work was drawn specifically from a customer service perspective but Gustafsson and Johnson (2003) suggest that this split between products and services is too simplistic, with most products offering a combination of goods and service. Berry et al (2002: 86) go further, claiming that 'offering products or services alone is no longer enough, companies must provide their customers with satisfactory experiences'. That is, from the consumer perspective, the experience is comprised of both the product and service, and separating the two has become an increasingly theoretical exercise.

The term 'first moment of truth' was coined by Normann (1984) with regards to the point at which the consumer first interacts with the product; with the packaging being seen as the main source of information. Once a product has been bought, or a service hired, then the interaction between consumer and business is continued and all further impressions made at this stage provide the 'second moment of truth' (Normann 1984). However, while the roles of advertising and marketing in the consumer / brand owner relationship are explored in theories on the consumer decision making process (see section 1.1.2), they are absent from both Normann and Carlzon's work. This suggests a more naive

consumer who only engages with a product or service provider at the point of purchase, rather than being exposed to them beforehand.

While there are other factors that have an impact on the first moment of truth, Lofgren (2005: 109) argues that ‘the first moment of truth deals with the ability of the packaging to grab the attention of customers’ and Underwood and Klein (2002: 59) suggest that ‘for decisions made at the point of purchase, packaging takes on a heightened importance relative to other communication tools – because of its easy availability’. Semenik (2002) also suggests that there has been a reduction in spending on traditional brand-building mass media advertising and a commensurate increase in point of purchase marketing materials. The role that packaging can play in the forming of relationships and in the decision making process of consumers, is central to the aims of this research. For many products, particularly those in the Fast Moving Consumer Goods (FMCG) category, packaging has been recognized as a critical element of brand differentiation and identity (Swientek, 2001; Bertrand 2002; Doyle 2002). While it is clear that additional factors may also affect the relationship between consumer and product at the first moment of truth, it is also argued that packaging can play a significant role. It is important, therefore, to explore ways in which greater insights into this relationship between consumer and product packaging can be elicited and analysed. Before looking at some of the factors that determine whether or not consumers are attracted to a particular product, it is important to understand why an individual might be looking to purchase anything at all.

1.1.1 A Wanting Animal

Maslow (1970) detailed a ‘hierarchy of human needs’ and described humans as a ‘wanting animal’. This means that we rarely reach a level of complete satisfaction and that even when we do ‘...it will usually only be temporary because once one desire has been fulfilled another will soon surface to take its place’ (Jordan 2000:5). As soon as needs at the bottom of the hierarchy have been met, attention will turn to meeting those higher up.

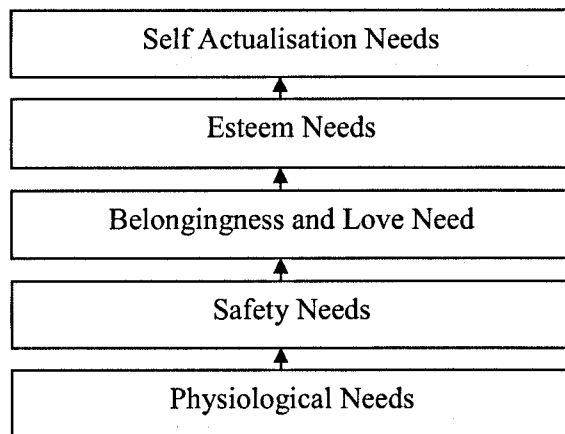


Figure 1 Maslow's hierarchy of needs (Maslow 1970)

In terms of consumer needs this has been adapted by Jordan (2000), who argues that the starting point is the consumer's functional needs, followed by usability needs and finally their pleasure needs, as displayed in Figure 2. This means that when consumers start to take it for granted that products will be functional, they then begin to expect products to be easy to use and also to provide emotional benefits.

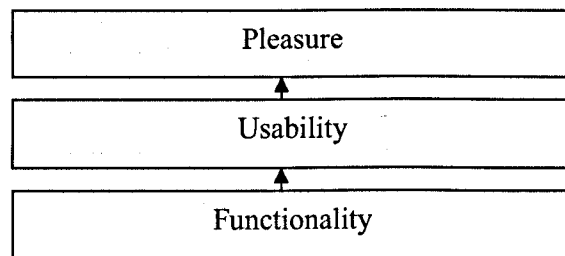


Figure 2 A hierarchy of consumer needs (Jordan 2000: 6)

Van Kleef et al (2005: 183) claim that 'the central goal in New Product Development is to create a product with superior consumer value so that consumer needs will be satisfied.' Functionality is usually the driving force behind a product. If the product is unable to do the job for which it is intended then it is unlikely to meet the demands of the consumer. Norman (2004: 37) suggests that 'a product's function(ality) specifies what activity it supports, what it is meant to do – if the functions are inadequate or of no interest, the product is of little value.' While consumers don't necessarily think about the

products that they use in these terms all of the time, the following examples from this research's data show that they are considerations that they can articulate.

1. *It's a different shape and more sort of practical like you say functional.*
2. *You'd expect it to be quite a powerful spray with something that size I would think.*

This suggests that not only do products need to have good functionality; they also need to *look* like they will work in the way in which the user is anticipating. If a product works well but looks like it might not, this could also have an impact on consumers' attitudes towards it. Once this base need of functionality has been met, the consumer then wants the product to meet their usability needs. Jordan (2000: 5) simply defines usability as being '...a product which is easy to use.' Again, the following examples from the data show that usability is something that consumers are aware of:

3. *I couldn't use it – I can't even get my hands round it properly to hold it.*
4. *But it's how practical it is how whether it would stand up or not.*
5. *You know some of these slip through your hands.*

Jordan also suggests that consumers are no longer impressed when something has good usability but are quickly unimpressed when it does not have it. This means that it has 'moved from being what marketing professionals call a 'satisfier' to being a 'dissatisfier'' (Jordan 2000: 3). Once a product has met these first two needs, it is then important that it meets the emotional or pleasure needs of the consumer. In line with Maslow's theory, once a consumer has achieved satisfaction on one level, they will automatically look to the next stage. This means that in addition to expecting our products to do their job and to be easy to use, we expect something extra from them. Jordan's work suggests that not only do we want our bottle opener to work and to be easy to use, we also want to derive some pleasure from using it and to have an emotional bond with it. As each need is met, the consumer is gaining a higher level of satisfaction from the product and is therefore

more likely to make a repeat purchase or to look at other products from the same company.

While Jordan's work goes some way to explaining the relationship between the different needs of functionality, usability and pleasure, this also raises further questions. The relative importance of these levels *between* similar products is left unexplored. For example it is unclear whether a bottle opener that is enjoyable, but hard to use, would be favoured over one that was easier, but less enjoyable.

Having discussed why an individual might want to buy a product, it is also important to understand how that decision process is carried out.

1.1.2 The Consumer Decision Making Process

From a brand owner's perspective, it is important to understand how a consumer chooses between different products, as knowledge of this process enables them to maximise their chances of being successful. Blackwell et al. (2005) propose a seven stage process for consumer decision making (see Figure 1.3). Their model has much in common with other proposed models of buyer behaviour (e.g. Nicosia 1966; Kardes et al. 1993 and Assael 2004) and, like Maslow, they suggest that the starting point for any purchasing decision is a need.

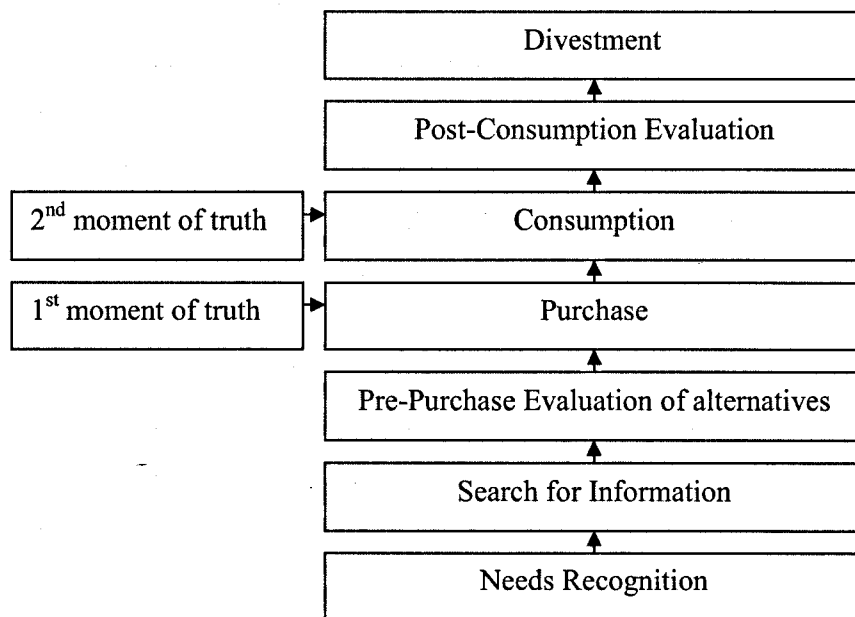


Figure 3 The Consumer Decision Making Process (Blackwell et al. 2005)

Figure 3 displays how the process starts with a need and then works its way through a Search, Evaluation, Purchase, Consumption, Post-Consumption Evaluation and finally the Divestment. Blackwell et al (2005: 72) define this need recognition as occurring when ‘an individual senses a difference between what he or she perceives to be the ideal versus the actual state of affairs’. The second stage incorporates a search for information about the product they want to buy. This can be both internal and external. Internal information would include drawing on prior knowledge and memories while external information could be gained from advertising material, trade magazines or talking to friends. It is here that the advertising and marketing messaging, which is largely absent from research relating to the first moment of truth, is accounted for in the consumer decision making process. However, while it is important to acknowledge the impact that these factors can have on the consumer, they lie outside of the scope of this research and for this reason are not addressed any further.

The third stage is where the consumer considers the different choices that are available to them in terms of the shop, the brand and the actual product. At this stage the consumer is evaluating both what have been termed ‘salient’ and ‘determinant’ variables (Blackwell

et al 2005). Salient variables such as cost and reliability are related to functionality and usability, as discussed above, and determinant variables are linked to emotional or pleasure needs. Each of these three stages takes place before the first moment of truth (Normann 1984) and therefore it is the fourth, purchasing stage, which is most relevant to this study. However the third, pre-purchase evaluation stage is important as it suggests that products are not just looked at in isolation but that consumers use a comparative approach when deciding what to buy. In addition to making inter-product comparisons, it is also possible that they make comparisons with an internal 'ideal' image of what a product from a certain category should look like. How these categorisations and comparisons are formed is discussed in greater detail in the introduction and discussion of Prototype Theory in section 4.3.5.3.

It is at the fourth, purchasing, stage that the consumer arrives at the store with an idea of the product that they want to buy. It is here that their decision making can potentially be affected by the brand owner through the packaging and design of their product and this is something that consumers are more aware of, as evidenced in the following comments:

6. *That seems to take away from the product that they're trying to disguise the product by fancy packaging.*
7. *I suppose the packaging could make it look more attractive.*
8. *You'd have your own packaging underneath to hold it up and look more marketable.*

However, in addition to marketing and advertising messaging, there are also other factors that can affect the consumer decision making process. Pham (2004: 362) argues that 'people use the valence of their feelings to infer the direction of their attitudes and preferences. If I feel good about something, I must like it; if I feel bad, I must not like it.' In other words, they look, in part, to how they are feeling, to assess how their attitudes about the product that they are contemplating buying. In addition to this, anything external that affects the consumers' level of arousal will also, therefore, affect their decision making process. The level of arousal that consumers have at the point of the

decision making will not remain constant. Instead it is likely to be in a state of flux dependent on their current emotions. Gorn, Pham and Sin's research (2001) demonstrates this effect in a study of advertising evaluations. In this study, music was used to manipulate the participants' incidental mood. Participants were then asked to evaluate an advert whose affective tone was either pleasant or unpleasant. They state that:

'When the target ad's affective tone was ambiguous, ad evaluation was more favorable among participants in a pleasant affective state than among participants in an unpleasant affective state.'

(Gorn, Pham, and Sin 2001: 47)

As one of many factors that can affect the consumer decision making process, the way in which underlying emotions are manipulated is discussed below in section 1.1.6. The Consumption and Post-Consumption Evaluation stages are related to the second of moment of truth where the product is being used and the consumer assesses their satisfaction, or otherwise, with the product. In addition to understanding the reasons why a consumer might choose a product and how they make their decisions, it is also necessary to explore what it is that they actually see when they are looking at such products.

1.1.3 Products as Attributes

Consumer behaviour theories (e.g. Engel, Blackwell and Miniard, 1994; Howard and Sheth, 1998) suggest that a product is conceived as a bundle of attributes which are the features that the consumer values as important. This means that when a consumer sees a bottle of beer or a soap dispenser, they are not simply looking at the overall design of the product packaging. Instead they are making many different judgments about each part or attribute of the packaging and making their decision about purchasing the product based on many or all of these attributes. Pham (1998, 2004) agrees that consumers view products as attributes but argues that this is too simplistic an approach as it does not take emotions or feelings into account, which can change which attributes, are seen as

important. The idea that products are viewed as being made of constituent parts is also supported by the following verbatims from the data:

9. *You know particularly the **grips** part of it.*
10. *I'm slightly concerned that the **spray mechanism** is exposed.*
11. *The **top part** there I like.*

While this was not something that respondents were asked about explicitly, the fact that they refer to the 'grips', the 'spray mechanism' and the 'top part' suggests that consumers do differentiate between different parts of the pack, rather than seeing it as a single piece.

Antonides (1991) suggests that which features are deemed important is determined by the consumer's intrinsic and extrinsic motivation. Intrinsic motivation is related to the demands and expectations of the consumer; they are likely to assess the available products in relation to their own requirements and uses and see how well each one compares. Extrinsic motivation is related to alternative products that they could choose from and suggests that in addition to comparing the product against their own expectations; they also compare products against each other to help assess which one will be most suitable for them to purchase.

Beck-Larsen and Nielsen (1999) suggest that low involvement products involve a relatively small number of concrete attributes, in comparison to more highly involved products. They describe low involvement products as those that are not likely to be of great importance to the consumer and write that there may be little difference between competitor brands or products. In their studies they use the example of vegetable oil as a low involvement product and a mobile phone as a high involvement product. If a consumer usually uses a particular type of vegetable oil but on one occasion that one is unavailable, then they would be likely to buy a different one. However, if a consumer was looking to buy a specific mobile phone and the shop they went to was out of stock, they would be likely to go elsewhere or come back at a later date, rather than simply buying an alternative phone.

Having argued that consumers do view products as sets of attributes, it is necessary to try and understand how these attributes are viewed in relation to each other. The following questions are central:

- Do all attributes have the same impact on consumer choice?
- If not, is it possible to tell which ones are most important?
- Do different social groups differ in relation to the importance they place on certain attributes?

Beck-Larsen and Nielsen (1999) argue that some attributes have a bigger impact on the opinions and behaviour of the consumer than others. For low involvement products the importance of each attribute is related to the cognitive effort the consumer makes in eliciting the attribute. Fazio et al (2000) suggest that the smaller the effort, the greater the influence on consumer choice and therefore 'top of mind' attributes are the most important with regards to consumer choice. Top of mind attributes are described as those which can be processed quickly and are immediately obvious to the viewer. If a possible important attribute was whether or not the product was made from recycled materials, this would only play a part in affecting the attitude of the consumer if it was quickly obvious that the product was made from recycled materials. If the recycling symbol (see figure 4) was obscured or hidden on the bottom of the product then this would be far less effective than if it was immediately visible.



Figure 4 The standard symbol to show that a product is made from recycled materials

Rajagopal and Burnkrant (2009: 238) concur with this view suggesting that the more accessible the attribute, the more likely it will be to determine consumer attitudes. 'Attributes that are recalled earlier from memory have a stronger impact on product attitudes than do attributes that are recalled later.' This further highlights the importance

of making differentiating attributes obvious and accessible to the consumer. Fishbein and Ajzen (1980) also suggest that the five to eight attributes elicited first are the most salient and important with regards to consumer attitudes and behaviour. Attributes mentioned after this are likely to be a result of deeper thought processes that are unlikely to take place when consumers are buying low involvement products.

Rajagopal and Burnkrant's (2009) work on hybrid products also highlights another factor which can determine the relative importance of different attributes. They describe hybrid products as those 'that possess features of multiple categories and therefore can potentially be categorized into more than one category' (Rajagopal and Burnkrant, 2009: 232). Perhaps the best example of these is the rapid advances in new product development within the mobile phone industry with many doubling as cameras, music players, personal organizers and so on. Apple's iPhone, which along with the many apps that it can run takes this to the extreme as it has the potential to function as an almost limitless number of products from across many different categories. This is particularly relevant given that the inspiration for much new product development research has its roots in the combining of products from different categories (Choy et al 2007). Rajagopal and Burnkrant (2009: 236) claim that 'the inconsistent attributes that separate the subtype from the main category become more important during evaluations than do the category consistent attributes'. The notions of categories and subtypes are explored in more detail in section 4.3.5.3 on Prototype Theory, but briefly this means that where a product is different from consumer expectations, it is the attributes that indicate the difference that will have a greater impact.

It has been argued that:

- Products are conceived as bunches of attributes.
- Some attributes are more important than others.
- These attributes affect the desirability of the product.

This would seem to suggest that it would be possible to discover which attributes are seen as most important for a particular product and then design these attributes to meet the aesthetic demands of consumers. However, although this would appear to be a theoretically sound approach to product design, there are two reasons why, even with a valid methodology, it is far from straightforward. The main problem is what has become known as the ‘attitude-behaviour problem’.

1.1.4 The Attitude-Behaviour Problem

The attitude – behaviour problem is the idea that although a respondent may indicate a certain attitude in speech, their actual behaviour in the real world might be different due to external factors that cause them to behave in a way that goes against their stated behaviour. This means that, a verbal or written statement is only ‘...a behavioural indicator of an attitude and there is not always a clear match between behavioural indicators and actual behaviour’ (Procter 2001: 106). This means that no matter how carefully the experiment is designed and how rigorously the analytical framework is applied, it is difficult to assess to what extent there will be a match between reported answers and subsequent behaviours. As with the following examples, taken from the data from this study, it is not possible to assess the validity of these comments.

12. *I'd buy that I'd buy that cos I quite like it.*
13. *Even if it were a little bit dearer I'd buy it.*
14. *No it would put me off well I wouldn't buy it anyway.*
15. *I probably wouldn't go out and buy it just for that.*

In terms of assessing the affective influence of a packaging solution, for a particular pack, this means that it is difficult to know if favourable reporting will result in the consumer being any more, or less likely to buy the product. Even if a respondent reports that a product is stylish, fun and desirable, it does not necessarily follow that they will purchase that product. This is all the more true for research in the early stages of the new product

development process where it is concepts or prototypes that are being tested, and as such there is no product in the real world to buy, or not. In addition, other factors such as price, size and availability come into play, as well as a multitude of other less stable factors such as customer's mood or state of mind at the point of decision of purchase. 'Many instances of human behaviour are "overdetermined" in the sense that a multitude of factors combine to produce them' (Ajzen 1988: 46).

Ajzen (1988: 4) describes an attitude as '...a disposition to respond favourably or unfavourably to an object, person, institution or event'; he goes on to caution that '...attitude is a hypothetical construct that, being inaccessible to direct observation, must be inferred from measurable responses'. However as discussed in section 1.2 below, the most common data type elicited within the mainstream qualitative consumer insight industry is consumer responses from a research interview setting. This type of data is not usually measurable in a quantifiable way and as will be discussed, a content or thematic analysis may be the extent to which it is analysed.

Indeed Ajzen's view would appear to suggest that attitudes can only be explored through quantitative means; even if what is being quantified is collected through more qualitative methodologies. As will become clear throughout this thesis, this is similar to the empirical approach taken in this work and this is discussed in detail in Chapter 5. However, while one of the central tenets of this thesis is that a combination of qualitative and quantitative techniques can add to existing approaches to analysing evaluative talk, I stop short of suggesting that a purely qualitative approach is inappropriate or unlikely to be successful in drawing out relevant insights.

1.1.5 The Theory of Planned Behaviour

The Theory of Planned Behaviour was developed by Ajzen (1988, 1991) and has since become the dominant account of the relationship between cognition and behaviour in social psychology and 'the most researched of these models' (Armitage & Conner,

2001: 47). Ajzen (2002; 665) argues that ‘human behavior is guided by three kinds of considerations’ and that the primary determinant of an individual’s behaviour is their intention to perform it. ‘Given a sufficient degree of actual control over the behavior, people are expected to carry out their intentions when the opportunity arises.’ (Ajzen 2002: 665)

The three considerations guiding behaviour are Normative, Behavioural and Control beliefs. Normative beliefs relate to the expectations of other people judging the behaviour I.e. peer pressure, and build a sense of a Subjective Norm. Behavioural beliefs are allied to the likely consequences of the behaviour and produce a favourable or unfavourable Attitude. Finally, Control reflects perceptions of the ease or difficulty of performing the behaviour, which in turn creates a Perceived Behavioural Control (PBC) The overall impact of these consideration is to moderate the intention to carry out a behaviour and ‘intention is thus assumed to be the immediate antecedent of behavior’. (Ajzen 2002: 665)

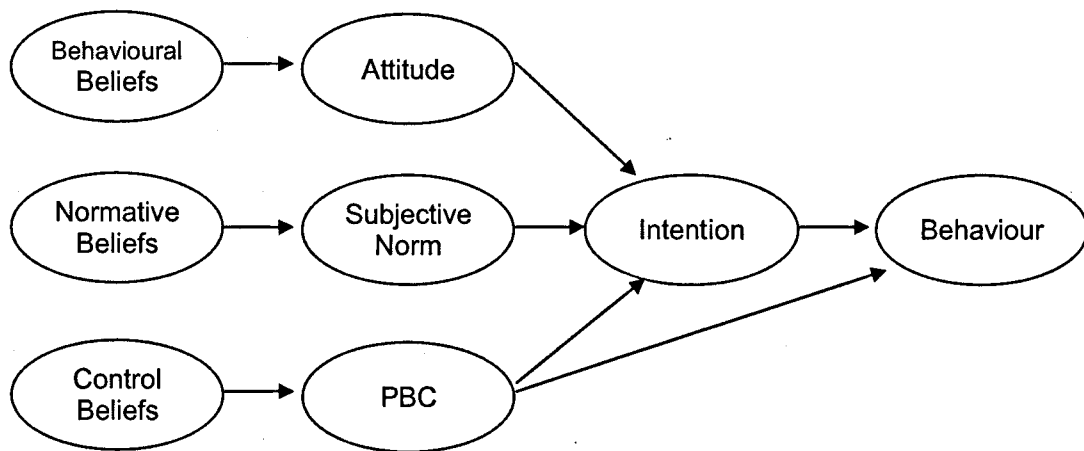


Figure 5 Theory of Planned Behaviour (Ajzen 1991)

Armitage and Conner (2001: 472) suggest that the ‘PBC provides information about the potential constraints on action as perceived by the actor, and is held to explain why intentions do not always predict behaviour’. In other words, if a behaviour is very easy to do, has high peer pressure and has positive, or low negative values attached to it, then this

behaviour may be carried out even if there was little intention. Perhaps the best example of this can be seen with Hornik et al's work (2001) which explored children's intent to take illicit drugs. Few if any declare any intent but for those who are subsequently exposed to a context where it is readily available, gains a high credibility cache and is seen as having positive outcomes, the taking of illegal drugs can become the norm, despite little previous intention to carry out the behaviour.

So while attitudes are still relevant, they are not viewed as being the single determining factor for a person's behaviour. It is also important to recognise that the relative weights of Attitude, PCB and Subjective Norms are not equal. Ajzen and Fishbein (2005: 195) attest that 'these weights vary as a function of the particular behavior and the population under consideration.' In general, individuals are also more disposed towards engaging in behaviours that they believe are achievable (Bandura 1997).

An understanding of the attitude behaviour relationship is one of the key issues for a research methodology reliant on the measurement of attitudes and evaluations. If there is little or no link between attitude and behaviour, then it may be ineffective to design products on the basis of consumer's reported attitudes. If the link between attitude and behaviour is not particularly strong then designs informed by such reported attitudes might not lead to better sales.

While there is an ever increasing understanding of the relationship between attitudes and behaviour, as a field, it has developed only relatively recently. Even towards the final quarter of the twentieth century, the prevailing assumption was that there was a one-to-one fit between attitudes and behaviour. A few early papers suggested that this view was far too simplistic (LaPiere 1934; Corey 1937) but it was not until Wicker published two papers arguing firstly that the link was typically not there, and then that it might be desirable to abandon the attitude concept (Wicker 1969, 1971) that the relationship was rigorously researched. Initially this led to the question: 'Is there a relationship between attitudes and subsequent behaviour?' While it became apparent that there were certain instances where there was a minimal or zero relationship (LaPiere 1934; Corey 1937), it

was also clear that strong correlations were also being observed (Kelley and Mirer 1974). These large differences in correlation led to a change in the line of questioning. Rather than asking if there was a fit, it became more relevant to ask when there was a closer relationship. 'Under what conditions do what kinds of attitudes held by what kinds of individuals predict what kinds of behaviour?' (Fazio and Zanna, 1981: 165).

The answer to this question is still not fully understood but Krosnick and Petty (1995) suggest seven cognitive variables that can potentially improve attitude-behaviour consistency. These variables are often inter-linked and correlated with each other, although research also shows that they can each be implemented in their own right (Cooke and Sheeran 2004). In relation to understanding the affective impact of a product, several of these variables seem particularly relevant. The level to which the respondent has a vested interest in the product has a definite impact on the correlation (Sivacek and Crano 1982). This suggests that if a consumer is giving an opinion about a product that they would be likely to buy, then the correlation would be higher than if it was for a product that was of little relevance to them.

This has obvious implications for the sampling of participants in this study. The fact that their responses are likely to influence the final design also gives them more of a vested interest in giving their attitudes. A second area of relevant work seems to suggest that the attitude-behaviour correlation is higher when people think in greater depth about their attitudes and have direct experience of the action that is the focus of the research.

'Findings show that direct experience tends to strengthen cognition-behaviour consistency' (Cooke and Sheeran 2004: 4). Conversely this means that the relationship is lower when people are less involved in the decision making. Current approaches to recruiting respondents cater to this relationship by ensuring that a screener is used, part of which determines the existing level of affinity that the respondent has with the brand or product category that is going to be discussed. The second point suggests that the research methods used should perhaps be those where the participant is given more time to consider their attitudes. However, using a more deliberative approach would mean that the research environment was different from the likely buying environment. As the

products and designs that are likely to be used in this research are low involvement, the consistency between attitude and behaviour may therefore be lower.

One factor that has perhaps led to a lower reported correlation between attitude and behaviour is the conflation of behaviours and goals (Ajzen & Fishbein 2005: 191) 'Some of the low correlations between intentions and behavior reported in the literature may occur when investigators try to predict a criterion over which people have relatively little volitional control'. They exemplify this with the 'behaviour' of losing weight. In this instance volitional behaviours such as going to the gym or having a healthier diet can be carried out, but weight loss may depend on physiological factors outside of the individual's control. Volitional behaviours can be understood as ones where the outcome is, within reason, under the control of the individual. Examples of volitional behaviours are voting in a general election, watching the evening news and donating blood. Ajzen (1988) suggests that 'When dealing with volitional behaviour people can be expected to do what they intend to do. Expressions of behavioural intention should thus permit highly accurate prediction of corresponding volitional action' (Ajzen 1988: 113). That the focus of this research can be considered a volitional behaviour, the purchasing of goods from a store, should increase the link between attitude and behaviour.

However, even if there is a strong correlation between attitude and behaviour, there is not necessarily an intrinsic link between attributes which are instrumental in the satisfaction of consumer needs, and therefore create positive attitudes, and consumer choice. For this to be the case, the products in the choice set have to differ on the attribute in question. If softness is an important attribute but all the products are made from the same material, and have equal softness, then this will not be the attribute that determines the choice. In real terms this means that not only does a producer have to work out what the salient attributes are for a particular product, they also have to make sure that their product has more of or is better than their competitors, in comparison to this attribute.

This section has provided an overview of the literature on the relationship between attitude and behaviour. However, as discussed below, consumer insight research of the

kind carried out in this study is not necessarily aiming to understand the behaviours, or intended behaviours, of the small sample used in the qualitative research. Rather, it aims to draw out insights from the evaluations of respondents with the intention of using these insights to inform the design of the new products. I have also explored the ways in which consumers arrive at the decision that they need to purchase a product and the ways in which they make that decision and view products. It is also necessary to try and understand how these experiences affect consumers on an emotional level and this is something that will be discussed below.

1.1.6 Manipulating Consumer Feelings

The notion that the immediate environment has an impact on individual's emotions is widely accepted in psychology and as Pham (2004) suggests, it follows that this emotional impact generally guides any subsequent interactions within that environment. Researchers have explored the relationship between emotions and specific products (Richins 1997) and also between the consumption of products and emotions (Bosmans and Baumgartner 2005; Garg et al 2005; Kidwell et al 2008). However while the existence of such a relationship is largely uncontested, Kidwell et al (2008: 158) argue that 'research has yet to fully understand how consumers use emotional information to make effective decisions.'

While there is much debate surrounding the specifics of the relationship, an understanding that emotions have an impact on consumer attitudes to products, and subsequently on consumer behaviour, has inevitably lead to interest from brand owners. Though perhaps questionable from a moral standpoint, the ability to manipulate consumers' emotions and therefore actions, at the point of purchase, is in some regards a natural extension of certain aspects of advertising and marketing communications. Research in the field of consumer manipulation has become a significant body of work under the term 'Atmospherics' (Areni and Kim 1993, 1994; Mitchell et al 1995; Hui et al, 1997 and Turley and Milliman 2000).

However, for many brand owners, products are predominantly sold in an environment which is out of their control. If a product is sold in a supermarket or local shop then the brand owner has little influence over the shopping environment or details such as the colour of the shelving, music played in the background or temperature of the room. This means that the only area of control that they have over impressions formed at the first moment of truth, is with regards to the first thing that the consumer sees of the product, that is, the packaging.

Having looked at some of the ways in which consumers interact with products, and the factors that ultimately influence their purchasing decisions, section 1.2 will discuss the way in which commercial researchers have tried to understand and measure these interactions.

1.2 Market Research

Although many different approaches and techniques have been used in the field of market research, its present day incarnation can be traced back to industrial pre-war America (Flick 2002). However, the style of this initial research was largely quantitative and resembled what would probably now be considered an audit or stock-take (Hague, Hague & Morgan 2004). Early methods centred on monitoring the actions of the consumer, rather than attempting to understand them. Whilst use of these methods grew in popularity, awareness of the importance of understanding the consumer was less prevalent. King and McDonald (1996: 54) point out that understanding consumer needs, rather than behaviour, came relatively late.

'In 1957 as markets moved from seller to buyer, new ideas of 'marketing' were taking hold [and] companies began to grasp the importance of understanding what the consumers really wanted; it could no longer be assumed to be the same as it always had.'

The impact of this change in the market is viewed by Kvale (1996: 71) as being one of the main contributory factors in the rise of qualitative market research. He suggested that 'In a consumer society, an extensive knowledge of the experiences, meanings, feelings, desires, and lifestyles of the consumer is essential to the design and marketing of consumer products'. Jaworski and Kohli (1993) suggest that it is only in recent years that understanding consumers has become an important research area that has made a positive impact on the success of both overall business performance and also the success, or otherwise, of many new products.

1.2.1 Market Research and Qualitative Interviews in the Development Process

Traditionally, market research has been used to assess consumers' attitudes to existing products rather than in the development of new products. As Valentine and Gordon (2000) suggest, the prevailing view has often been that consumers do not really know what they want, with the assumption being that consumers are passive recipients of the business' decisions and judgements. However, as the advantages of creating a dialogue with the consumer in other areas of market research have become apparent, this same process has now spread into the field of New Product Development. Bogue and Ritson (2006) highlight the need to utilise the end user in the development process within the food industry, and Narver and Slater's research showed that 'firms who adopted a market orientation achieved higher levels of relative profitability, sales growth and new product success' (Narver and Slater 1990: 29). Johnes (1994: 52) summarises this change in the collective viewpoint concluding that 'there is little disagreement in the literature that listening to the voice of the market is important for product development purposes.' A strong body of work has developed in support of this assertion (Foxall & Johnston 1987; von Hippel 1988) and the advantages of using market research techniques to aid the development of new products are becoming clearer.

However, exploring the consumer view is still not considered essential within all stages of this process. Van Kleef et al (2005: 181) suggest that 'it is increasingly recognised that

successful new product development strongly depends on the quality of the opportunity identification stage... [but]... that asking consumers what they want is useless, because they do not know what they want'. In other words, while it is becoming clearer that consumers should be consulted once a product is in the process of being developed, it is still less common for consumers to be questioned with regards to gaps in the market for such potential new products. In addition, there are still many products and aspects of products that are designed by 'experts' and not explored with consumers at any stage of the development process.

Hague, Hague and Morgan (2004: 61) give a more detailed account of how the consumer might be involved in several stages of the developmental process, including how they can be used to identify new opportunities:

'A survey on product development might begin with focus groups to explore unmet needs, followed by structured interviews to measure the size of these needs, and conclude with depth interviewing to test the concepts.'

As can be seen in section 5.2, this is similar in kind to the empirical approach that was used in the main study of this thesis.

The term 'Market Research' encompasses many different approaches to eliciting information from consumers. Although there has been a certain merging of the two areas (Smith & Fletcher 2004), it is still a field broadly split into two camps; qualitative and quantitative research methods. A more detailed account of the primary differences between the two is offered by Silverman's book on research methods (2006: 33-63), but the following table provides an overview of some of their claimed differences.

Qualitative	Quantitative
Soft	Hard
Flexible	Fixed
Subjective	Objective
Political	Value free
Case study	Survey
Speculative	Hypothesis testing
Grounded	Abstract

Table 1 Claimed features of Qualitative and Quantitative methods (Halfpenny 1979: 799)

It should be immediately obvious, here, that qualitative research has been negatively positioned. Denzin and Lincoln (1994: 4) state that the work of qualitative researchers has historically been viewed as ‘unscientific, or only exploratory, or entirely personal and full of bias.’ While this view is no longer so prevalent, indeed Kvale (1996: 71) points out that ‘qualitative interviews are extensively used in today’s market research’, Patton (2002) argues that there is still the prevailing view that qualitative research is somehow less scientific or replicable than its quantitative sibling. Despite this, it is still evident that for many research projects, a qualitative approach that is more flexible and detailed in both its design and analysis will provide a greater level of understanding of consumer opinions than a larger scale but less personal, quantitative approach.

Having argued that consumer focused market research is now being viewed as a useful tool within the new product development process; I will now briefly introduce the most prevalent elicitation approach. Although it is used in several different guises, the most common of these approaches is the qualitative interview. This is discussed in much more detail in section 4.1.1.2 but briefly, qualitative interviews can take the shape of a focus group, a one-to-one interview or a paired-depth interview. Once the interviews have been carried out, the next step is to use the information that has been collected. Drawing on expertise and knowledge from many different fields, the analysis of qualitative data draws on a range of different analytical techniques.

'Today there are a vast number of different research approaches drawn from academic and scientific disciplines that have made qualitative research an eclectic method of enquiry.'

(Valentine & Gordon 2000: 8).

Though many different analytical approaches are used, the use of linguistic analysis is still infrequent and this is discussed below.

1.2.2 Approaches to Analysing Qualitative Data

As stated in the introduction to this thesis, one of the potential applications of this research is with regards to the development of new methodologies for analysing consumer evaluations. In this section I argue that a linguistic analysis might be necessary as a more surface level, or intuitive analysis, is likely to miss certain features or aspects of an evaluation. A greater understanding of the underlying factors that have an affect on evaluative talk is therefore likely to be beneficial in a market research setting.

It has long been established in the linguistic literature that meaning cannot simply be read from the surface of a text. Both the fields of Critical Linguistics (for example Fowler et al 1979; Hodge & Kress 1979; Fairclough 1989, 1993, 1995) and Systemic Functional Grammar (Halliday 1994) are centred around the premise that there is as much, if not more, meaning contained in the way something is said, as there is in simply reading what has been said. Fowler et al (1979: 9) argue that 'Significance cannot simply be read off the linguistic forms [...] in the text.' Fairclough (1993: 28) concurs, stating that '...Meanings of discourse cannot simply be read from the text'. The aim of this research, therefore, is to gain insights into consumer attitudes that have previously proved elusive, possibly because they cannot be read directly from the text.

Although not drawn from a market research context, the point that meaning is often embedded within a text or between speakers is a salient one. However, it is only relatively recently that this knowledge has been applied to the context of qualitative data

analysis in market research. In his book on qualitative research interviews, Kvale (1996: 201) states that the analysis should go 'beyond what is directly said' in order to understand 'meaning not immediately apparent in a text'. While there seems to be an understanding that meaning often lies beneath the conversational surface, this hasn't necessarily translated into a widespread exploration of how these insights might be accessed.

'In most market research studies, the way language is used is clearly important. But this usually falls short of needing to conduct an etymological and/ or syntactical analysis of the way words have appeared, and been used. Commercial market researchers clearly need to take into account the way in which individuals will discuss brands taking care to examine the language that is deployed.'

(Smith & Fletcher 2004:101)

Smith & Fletcher (2004: 101) also highlight other common features of spoken language, such as false starts, repetitions and overlapping speech as being of potential interest, but caution that 'For most commercial research, it is generally accepted that the analyst will judiciously edit the verbatim comments to avoid unnecessary repetition and deviation.' In many cases it may be the elements that are edited out that are of particular interest to the linguist and therefore the market researcher.

On the one hand, then, is the tacit acknowledgement that there is additional information to be explored, but on the other it is accepted that for the most part this exploration is not carried out. This should not, perhaps, be so surprising. Flick concludes that the findings of social research have not found their way into institutional use as much as might be expected. He also point out that even where such methods are taken up, they are 'obviously reinterpreted and picked to pieces' (Flick 2009: 31).

There is much discussion in the literature on the changing views on the acceptability and suitability of intuition (De Vos 1998, Bryman 2001 and Smith and Fletcher 2004). It is

argued that just because an answer is borne out of intuition, or from a hunch, it does not necessarily mean that there is a complete absence of an organized process. Smith and Fletcher (2004: 8) suggest that there is currently a certain level of acceptance towards intuition but caution that ‘...the pendulum must not swing too far’ and that the key to success lies in the successful marriage of ‘...‘informed’ intuition with the rigorous scrutiny of data’. Smith and Fletcher (2004) also suggest that qualitative analytical approaches, which are based on an open and flexible way of thinking, can also be used to help interpret quantitative data. If this is the case, then the opposite should also be equally true. A more systematic and structured analytical approach could also be successfully applied to qualitative data and it is this approach that is taken for the analysis of the data within this study

The previous sections have established that consumers’ opinions about a product can be affected by its design. They have also shown that despite some limitations, there is a link between the consumer liking a product and then going on to purchase it. This knowledge suggests that testing products for their affective impact should be part of the research and development process. The knowledge that their new product or packaging has a positive affective impact could give producers a higher level of confidence with regards to its impact and success. The previous sections have also explored the reasons why a consumer might want a product, how they decide between products and, what it is that they actually see when they are looking at products. They also highlighted some of the problems associated with trying to make connections between intent and actions based on linguistic information elicited from consumers. Finally they have shown that qualitative, linguistics approaches are starting to be used within market research and new product development practices. Chapter 2 will now explore some of the ways in which language can be analysed to account for its evaluative content.

2. Interpreting Evaluative Language

This chapter explores approaches from four different levels of linguistics with regards to understanding and analysing evaluative language. It proposes that the structure of the communication can give an indication as to the encoded strength of the contents. Certain communicative constructs are identified as being more difficult to make for the speaker and it is therefore hypothesized that evaluations contained within these structures carry more importance than those made within structures that are easier to make. A scale of the inherent strength of an evaluation, based on the politeness and turn taking structures that are involved in the production of any given utterance is suggested. By combining this with an analysis of the more explicit evaluative content within a turn, it is argued that a deeper interpretation of evaluation can be achieved. The four levels that this chapter explores are:

- Discourse
- Grammar
- Lexis
- Phonology

Each of these areas is introduced and discussed with regards to how they could be utilized to develop a framework for evaluating evaluative language. Section 2.1 on discourse explores the way in which evaluation is created and can be interpreted as a social construct and concludes by exploring to what extent it is possible to develop a scale of inherent importance, as introduced above. Section 2.2 on grammar explores the different ways in which evaluations can be grammatically realised and section 2.3 on lexis introduces the concept of semantic prosody and evaluation being contained both within and between words. Finally section 2.4 introduces phonological and non verbal aspects of communication before dismissing them as suitable analytical approaches for this research

2.1 Discourse

For Fairclough (1993) the analysis of discourse can be split into two separate areas. Firstly there is the critical approach that has ‘a particular interest in language and power... [and] ... aims to investigate critically social inequality as it is expressed, signalled, constituted, legitimised and so on by language use.’ (Wodak & Meyer 2001: 2). Critical Discourse Analysis (CDA) is concerned with making explicit the hidden meanings and ideologies which are, it claims, encoded within all texts. ‘Language is an instrument of control... it involves the systematic distortion [of reality] in the service of class interest’. (Hodge & Kress 1979: 6) It is this interest in power relations that sets it aside from other branches of discourse analysis. However, because I am investigating language at the micro-level rather than in its larger social and political context, I am more interested in the second, non-critical approach, which is focused on the view that language is based on social interaction. It is concerned with the social contexts in which discourse is embedded and covers fields as varied as class room discourse (Sinclair and Coulthard, 1975; Thornborrow, 2002; Jones and Thornborrow, 2004), and the discourse of advertising (Myers, 1989; Cook 2001).

I argue that to gain a more complete understanding of evaluative talk it is necessary to account for the social context of interaction. This exploration also forms the basis for answering the second research question proposed in the introduction in Chapter 1:

‘What are the social and interactional considerations when looking to understand linguistic evaluations?’

The following section on Conversation Analysis details the way in which this social aspect of communication has previously been explored.

2.1.1 Conversation Analysis

Conversation Analysis (CA) is an empirical approach to the study of spoken language borne out of Schegloff (1968), Sacks (1972, 1974, 1984) and Jefferson's (1978) pioneering work in the field of ethnomethodology which itself built on the sociological approaches of Garfinkel and Goffman in the 1960s. Working in the field of sociology these scholars were interested in developing techniques that would allow them to better understand how people make sense of everyday life with Sacks (1984: 18) claiming that '... the detailed study of small phenomena may give an enormous understanding of the way humans do things'. Conversation analysts attempt to identify and understand patterns and structures that underpin talk in action. Goodwin and Heritage (1990: 283) suggest that CA places the emphasis on '...participants' orientation to indigenous social and cultural constructs' describing the underlying social organization 'through which orderly and intelligible social interaction is made possible'. One of the main aims of CA then, is to understand and describe the competences that ordinary speakers use and rely on in participating in meaningful and useful conversation. It is argued that meaning is constructed in situ by speakers who are:

'Simultaneously engaged in fine-grained real time co-ordination of speaking turns tracked predominantly in terms of surface structural features and... organizing their actions in terms of publicly accountable normative expectations bearing on the nature and design of their turns at talk' (Heritage 1989: 26)

The importance of embracing an analytical approach that looks at both the contents and the structure of a communicative exchange is argued by Goodwin and Goodwin (1992: 182) who wrote that 'particular interpretations of events in the world may be far less important than the structures used to accomplish such congruent interpretations as a social activity in the first place.' In other words, the roles of social and politeness structures should be as much a point of exploration as the actual contents of what has been said. Litosseliti (2003: 86) also suggests that 'how language is used is as important

as the words themselves' and Stewart, Shamdasani, and Rook (2007: 23) suggest that due to the composition of any group, speakers are likely to have an influence on each others' responses as '...personality characteristics interact with demographic variables to influence the behaviour of individuals in the group'.

In terms of consumer insight research and focus groups, Myers (2007: 81) argues that 'the way participants (and moderators) say things can be as important as what they say' both in terms of the content and also the impact on subsequent turns in the discussion. Other research exploring focus group interactions has presented similar arguments (Myers 1998, 2004; Matoesian and Coldren 2002; Puchta and Potter 2003; Puchta et al. 2004).

Despite this, exploring these interactions is not traditionally part of a focus group analysis and Morgan (1988:26) suggests that there is 'typically little attention to either the micro-dynamics of the interaction process or the contextual constraints of the focus group setting'. Litosseliti (2003) describes a range of analytical approaches from detailed content and thematic analysis based on transcripts through to a summary or report based on the researcher's 'continual analysis' but Myers (2004) argues that in a market research setting there is often little time or understanding of the need for full transcription or detailed content analysis, and so the latter is far more common.

In a review of more than forty studies using data drawn from focus groups, Kitzinger (1994) found that none of them concentrated on the conversations between the participants. Instead the focus was on the moderator to participant interactions, topic management, or other topics investigated by conversation analysts.

The concept of meaning in conversation being a joint production between speakers is introduced below. Four main areas within CA that are of interest to this research are then explored. Firstly, the idea of turn taking is discussed in section 2.1.1.2 with the concepts of preference and power being introduced in section 2.1.1.3. The relevance of Content and Non – Content turns is discussed in 2.1.1.4 and Politeness Theory is then also

introduced. Each of these can be understood in terms of the idea that all meaning within a conversation is a product not just of the individual who made the utterance, but of each of the participants involved. This section concludes by proposing a framework for exploring the inherent strength of evaluations based on both the structural and social elements that are involved in the production of any given utterance.

2.1.1.1 Meaning in Conversation as a Joint Production

Conversation Analysts also developed the idea that any given narrative or response is jointly produced by any present speakers and hearers. Fitch and Sanders (2005: 105) argue that:

'The meaning of an action is heavily shaped by the sequence of previous actions from which it emerges and that social context itself is a dynamically created thing that is expressed in and through the sequential organization of interaction'.

That is, the utterances of any speaker in the communication are likely to be, in part, determined by the presence and action of other co-locutors. This has been found across different genres including storytelling (Mishler 1995), research interviews (Jin 1992) and business negotiations (Lampi 1986). This notion is particularly relevant in a market research interview setting as it is important to minimise any moderator affects that might influence respondents' utterances. Understanding the way in which one speaker's utterances can influence the other's might also prove useful in trying to gain a greater level of insight into the evaluations that are being made. Although Wolfson (1976) found that research interviews lacked the 'performance features' that are indicative of evaluative communications, she later amended this to being dependant on the level of shared agreement between the participants. Cortazzi (1991) and Jin (1992) have both subsequently found examples of performed narratives in research interviews.

Cortazzi and Jin (2000: 110) propose that when a respondent speaks they are responding 'not just to interviewer questions but to an interviewer's assessments of previous answers and narratives'. This also has implications for the analysis of research interviews as it means that responses to previous answers might also be explored to see how they could have affected subsequent answers.

2.1.1.2 Turn Taking

The notion of turn-taking stems from Sacks et al's study on the organisation of turn taking (1974). They note that although the idea of turn taking had been explored earlier, 'no account of the systematics of the organisation of turn-taking for conversation is available' (Sacks et al 1974: 696). Using tape recordings of naturally occurring conversations, they set out to develop a system that they claimed was lacking. Using data from these recordings they observed fourteen key structure points (Sacks et al 1974: 700-01):

1. Speaker change occurs, or at least reoccurs;
2. Overwhelmingly, one party talks at a time;
3. Occurrences of more than one speaker at a time are common, but brief;
4. Transitions with no gap and no overlap are common. Together with transitions with slight gap or slight overlap, they make up the vast majority of transitions;
5. Turn order is not fixed, but varies;
6. Turn size is not fixed, but varies;
7. Length of conversation is not specified in advance;
8. What parties say is not specified in advance;
9. Relative distribution of turns is not specified in advance;
10. Number of parties can vary;
11. Talk can be continuous or discontinuous;
12. Turn allocation techniques are obviously used. A current speaker may select a next speaker or parties may self-select in starting to talk;

13. Various 'turn constructional units' are employed; e.g. turns can be projectively 'one word long' or they can be sentential in length, or longer;
14. Repair mechanisms exist for dealing with turn taking errors and violations.

Schegloff's study on 'Sequencing in Conversational Openings' was confined to what he termed 'one limited aspect of conversation' (1968: 1075) but this, and other work like it, was still useful in showing that raw data from real life conversations could be subjected to rigorous analysis. Other research, by Jefferson (1973) and Gunter (1974) also looked at specific areas that could be grouped under the heading of 'Turn-taking', without setting out to produce a comprehensive discussion and explanation of what these different features showed about the interlocutors and what they were saying. Although this early work in the field of CA is historically important in terms of the foundations it set out for the observation and analysis of language, its overall approach was structural rather than sociolinguistic. While some of CA's early observations listed above are more relevant to research exploring evaluative talk than others, particularly those that comment on interruptions and speaker selection, it is perhaps inevitable that these early approaches would be adapted to understand more about *why* interactions are structured in this way and what this could tell us. Despite not exploring *why* interaction is structured in such a way, this early work does suggest that biases within turns might exist with regards to determining the next speaker and what that might mean for the interpretation of the subsequent response.

Eggs and Slade (2001: 6) report that much early work in CA was in relation to adjacency pairs and the effect that they have on the structure of a conversation. Hutchby and Wooffitt (2004: 39) describe adjacency pairs as 'one of the most noticeable things about conversation' reporting that 'certain classes of utterance conventionally come in pairs. For instance questions and answers; greetings and return greetings; or invitations and acceptances/ declinations'. This structural approach is introduced below with subsequent and alternative areas of research then being discussed in relation to how they facilitate an exploration of the relationship between structure and evaluative strength.

The most common pattern in a two person communication is that of speaker A being the previous and next speaker to Speaker B and of Speaker B being the previous and next speaker to Speaker A (Lee 2007). In other words, they take alternate turns and nominate each other, directly or indirectly, to continue this pattern. The point of interest occurs when a third speaker either interrupts this pattern or is nominated by speaker A or B. This nomination often takes the form of a request for information or solidarity on a view point on which the other speaker might disagree. If the most prevalent pattern is for speakers to alternate turns, even when there are other participants present, then this will add to the potential cost of breaking this pattern. Therefore evaluations contained within turns that have this potential higher cost are likely to be stronger opinions than those that do not have these associated higher costs. This is particularly relevant when a speaker self selects their turn rather than being invited to take the floor. However, Sacks et al (1974: 709) quickly determine that 'the sources of this bias are external to the turn-taking system's basic organisation and cannot be detailed here'. It is precisely these kinds of areas that are of interest to this research, particularly in relation to answering research questions one and two. The associated cost related to a turn, and the differing implied strength of an evaluation, is one of the areas that this research seeks to develop.

As stated above, although two part exchanges were some of CA's earliest areas of exploration and continue to be studied, research in this field has since diversified to develop alternate models for understanding the structure of everyday communication. Sinclair and Coulthard's (1975, 1992) work on classroom discourse followed on from this earlier work and they set out to develop a tool for systematically studying the nature of interactions between teacher and pupil. Using classroom discourse they developed a system that has much in common with other systemic grammars in the way that the different units relate to each other. The Sinclair and Coulthard model is a rank scale model and consists of five ranks. These are *lesson*, *transaction*, *exchange*, *move* and *act*, and these are related to one another in a 'consists of' relationship' (Willis 1992: 112). The ranks are hierarchical in nature with *lesson* being the largest unit and *act* being the smallest. Sinclair and Coulthard identify twenty-one different classes of *act*, which combine to make the five classes of *move*. These are *framing* and *focusing moves*, which

combine to make *boundary exchanges* and *opening, responding* and *follow-up moves*, which combine to make *teaching exchanges*. A number of these *exchanges* combine to make *transactions*, which combine to make the *lesson*.

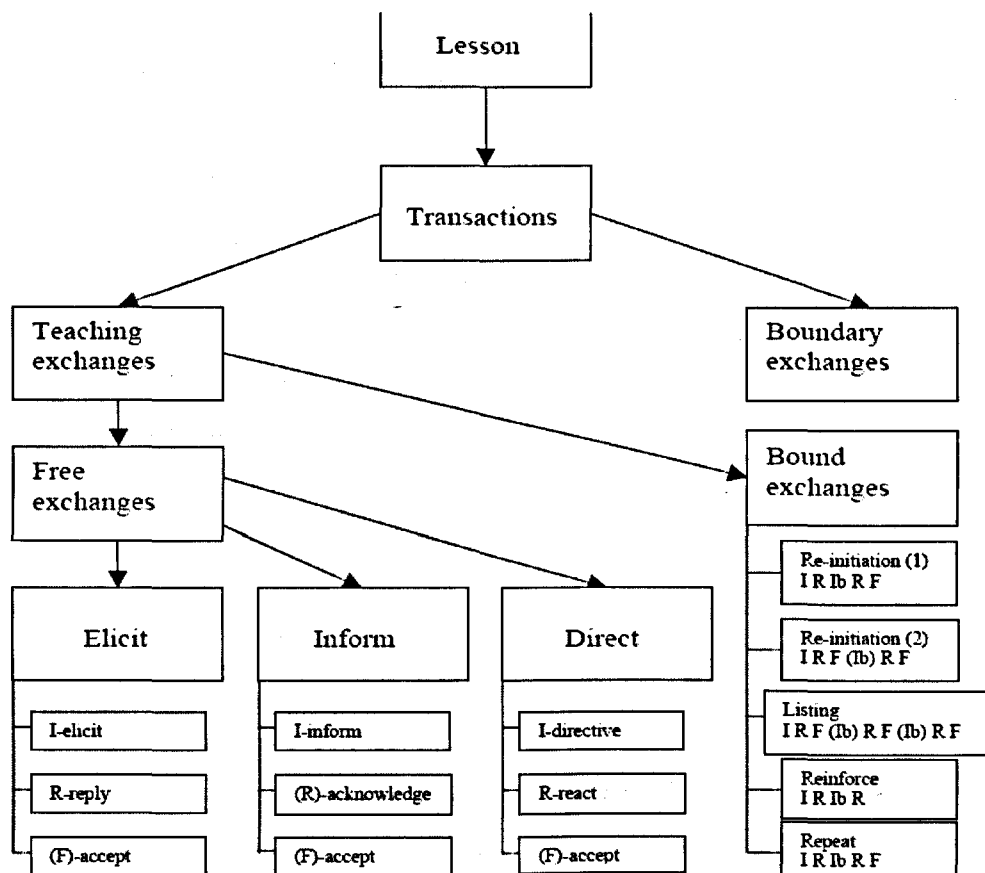


Figure 6 A diagrammatic representation of Sinclair and Coulthard's Initiation-Response (Farooq 1999: 31)

This shows the hierarchical nature of the model and some of the different categories available to the analyst. The basic move types of, answering, eliciting, informing, are borrowed from more traditional work on Speech Act Theory (Searle 1969). However Sinclair and Coulthard's work on Discourse Analysis differs from research in the Conversation Analysis tradition in that it eschews the two-part adjacency pair model proposed by Schegloff and Sacks (1973) in favour of a three part exchange model.

The primary structure is the Initiation-Response-Follow-up (IRF) exchange whereby an initial question, assessment or comment is made, a response comes from the second speaker and the first speaker then has a further turn to close the exchange. Lee (2007) suggests that this pattern is mostly found in classroom interactions between teacher and pupil but there also many examples from my data that follow this pattern.

16. Moderator: *Any other thoughts on this one*

Speaker A: *No I think that's it for that*

Moderator: *Ok no problem let's move on to the next one then*

Here the third turn closes the line of questioning for a specific stimulus and moves it onto the next item to be discussed.

The third response or evaluation turn is an area of much interest in terms of the influence it has on subsequent turns. Lee's study of the third turn in a three part exchange argues that the third turn can take many different forms and have many different roles depending on the context of the exchange and on the preceding turns. The teacher, or in the case of my research, the moderator, can respond not only to whether the second turn answers are 'correct, adequate or relevant but also to how they are produced: accurately, convincingly, or reluctantly' (Lee 2007: 1205). The notions of a correct response should be immaterial here as the questions asked by the moderator are genuine requests for opinion rather than a test of the participant's knowledge. However as Puchta et al. (2004: 290) state: 'In a focus group there is no right answer, although there may be the right kind of answer'.

The effect and role of different question types is discussed below. While the way in which the answer or evaluation or response is made might be of interest to the moderator, it is the adequacy or relevance that is more likely to be the subject of the third turn.

17. Moderator: *So erm what kind of person do you think that this would be who would this be aimed at if it came out in the shops?*

Speaker A: *More men I'd have thought.*

Moderator: *Ok right so what kind of age range or sort of what about it is more masculine then if that's what you think.*

In this example the participant's answer is both relevant and correct, in as much as it answers the question. However, the moderator feels that there is more information that can be retrieved from the participant so uses the third turn to ask a further question, and to explore some of the reasons behind the initial answer that was given. This then acts to evaluate the previous answer and to set off another three turn sequence following the pattern of alternating speakers.

Though this area that has been extensively researched, particularly in relation to teacher / pupil interactions, (Tapper 1996; Nassaji and Wells 2000; Macbeth 2003), Lee suggests that the norm of using formal categories and boundaries misses some of the point of the third turn. She contends that 'the formal categories take for granted, and often get in the way of making sense of, the contingent interpretive acts the third turn brings out and what is accomplished in the process' (Lee 2007: 1206). Drawing on data taken from forty-six hours worth of classroom interactions, she proposes that an alternative approach should be considered that 'focuses on cases and instances of interactional details, not necessarily made to fit the formal categories or the relations among them' (Lee 2007: 1226).

Sinclair and Coulthard's framework claims to account for the majority of teacher pupil interactions but it has still faced some criticism. Indeed Malouf (1995: 1) argues that Sinclair and Coulthard's model '...has only been applied to two-party discourse and would seem to fall short of the full range of linguistic communication.' Francis and Hunston (1992) also point out the importance of paralinguistic features and the role that they should perhaps have within discourse analysis. Classroom interactions and research interviews are similar in being semi-formal and having one dominant participant, the teacher or moderator, though they are of course different. So, while this framework is relevant here, Sinclair and Coulthard (1975, 1992) acknowledge that the system was not designed to handle pupil to pupil interactions or discussion groups - which can be seen as

analogous to my paired depth interviews. Therefore it would not be unexpected that some amendments might be required to allow the system to handle this different kind of interactional data.

When looking to establish the importance that should be placed on any given evaluation or turn, I argue that one consideration could be to look at the cost or difficulty of making that turn. One way of establishing that is to look at the way in which the speaker of the turn has been selected. Sacks, Schegloff, and Jefferson (1974) suggest that turn-allocation techniques are distributed into two groups; those in which next turn is allocated by current speaker selecting next speaker:

18. Speaker A: *Ok that's great John any feature of is that you do like at all.*
Speaker B: *The handgrips I do like the handgrips.*

In this example speaker B's turn has been allocated by speaker A through the use of direct address.

The second of these groups is those in which a next turn is allocated by self-selection;

19. Moderator: *Ok er and er do you think it would work for any particular brand at all or do you think it could work for all brands or.*
Speaker A: *It looks it looks like something from Sure to me do not look like Lynx*

Here the moderator asks an open question that gives both participants the opportunity to answer. Speaker A self selects their turn and Sacks, Schegloff, and Jefferson's model (1974: 707) suggest that one marker of this is the non-use of a 'speaker selects next' approach from the previous speaker.

A turn that has been directly solicited by the previous speaker is perhaps easier to make as it is obvious that the second speaker is being invited to talk and therefore there is less potential cost to their turn.

20. Moderator: *And er David what do you would you say that was more for men than women or.*

Speaker B: *Erm well it would have to be I mean there's no way a woman is going to get her hand round that.*

The direct address used by the moderator marks out who is expected to take the next turn right from the start of their own turn. However as Sacks, Schegloff, and Jefferson (1974: 705) point out, 'the accomplishment of turn-transfer does not occur until the first possible transition- relevance place.' Direct address isn't the only way of selecting the next speaker though as this can also be achieved through reference to previous topics or ideas related to a specific speaker and cues related to body language such as glance or pointing.

21. Speaker A: *Yea I'd say that was aimed at more female*

Moderator: *Ok we've said more female and how why do you think that*

Speaker A: *Just because it's smaller more compact like more for handbags or something*

Referring back to the 'femaleness' of the stimulus marks out the speaker who previously proposed this is being invited to continue on that theme. Although they have not been directed directly, the fact that the question is aimed at the topic on which they had previously commented provides a strong indication as to who is expected to take the conversational floor.

Having argued that a turn that comes in the form of an invited response has less potential cost to the speaker, this also has implications for the attached cost to an initiation turn. If the evaluation is unsolicited, that is, it is not in response to a direct request or following on from 'topic reference' then it perhaps has a greater inherent strength and is indicative

of a stronger opinion for several reasons. Firstly, this initiation turn indicates that the speaker feels strongly enough to provide the evaluation, without being asked as there is no structural imperative to provide one. Secondly it is likely to have been produced prior to the moderator asking any questions on that topic or discussion area and before a fellow respondent has commented. In this way it can be considered a 'clean' evaluation in that the effects of leading questions or (dis)preference structures are minimised. As discussed in section 1.1.3, the fact that the evaluation is an initial assessment is also indicative that it is a stronger opinion as it has shown that there is a direct correlation between the order in which judgments or assessments are made and the strength of that evaluation (Fishbein and Ajzen 1980, Fazio 1986). A respondent giving an initiation turn also has less information to go on in terms of whether the opinions that they are giving are 'right' and this type of doubt can be seen in comments such as *'Is this the sort of thing you were after?'*, *'Am I supposed to be saying things like that?'* In each of these utterances the respondents' conclusion of their turn is a check to see that what they have previously said is 'correct' in terms of being relevant and appropriate to the task.

2.1.1.3 Preference and Power

Two additional areas that have an impact on the co-production of speech are those of Preference and Power. These are both introduced and discussed in detail below with Preference relating to whether or not a response 'agrees' with the preceding turn and the section on Power exploring how power differences between speakers can impact on these preferred or dispreferred constructions.

2.1.1.3.1 Preference

Following on from observations relating to adjacency pairs, researchers in CA developed their field of study to account for what has been termed 'preference' (Sacks 1978; Pomerantz 1984). Atkinson and Heritage (1984: 53) contend that in conversation choices between courses of action are 'routinely implanted in ways that reflect an institutionalised ranking of alternatives.'

Pomerantz (1984: 54) argues that a first speaker's assessment of someone or something invites a subsequent assessment as a matter of course and that this can take the form of either an agreement or a disagreement. Hutchby and Wooffitt (2008: 46) state that 'the format for agreements is labelled the 'preferred' action turn shape and the disagreement format is called the 'dispreferred' action turn shape.' One of the most relevant and well researched patterns in conversation analysis is the desire for agreement or a preferred response in the second turn of an adjacency pair. That is, when one person asks a question, makes an assertion, or performs some other conversational action; they are likely to expect not only a response but also one that is to be taken as agreeing with them. The consumer insight setting of this research means that in turns between the moderator and a participant, there should not be preferred or dispreferred responses due to the nature of the questions. The type of question used by the moderator means that the responses should be neutral in terms of their (dis)preferred status. Long and Sato (1983) looked at the way teachers used different questions and found that there were two main types - referential and display. Lynch (1996: 108) describes a display question as one 'in which the teacher knows the answer in advance.' In other words, it is being asked to test the answerer not to actually try and find out an answer. This is of course the opposite to the aims of moderator questions in a consumer insight interview. Not only should the moderator avoid having any preconceived notions of what constitutes a 'correct' answer, they should make it clear to the participants that there are no wrong answers and they are simply looking for their opinions. So in this setting the questions are referential in that there is no correct answer and the moderator is genuinely asking the questions to find things out. This means that for a referential question there should not be a preferred or dispreferred response as the participant should have no idea what the moderator 'wants' them to say.

In addition to highlighting a preference for agreement between speakers within individual turn structures, research in conversation analysis has also shown a preference for consensus between speakers in general with Watt (2003: 50) arguing that 'friction in personal interaction is undesirable'. This is potentially problematic for research that aims

to explore a range of different attitudes and opinions with regards to ideas and objects. This preference for agreement may lead to participants avoiding giving dispreferred responses or even to avoid giving opinions that they think will be different to or disliked by others. Turns between participants are likely to have more obvious preferred or dispreferred response markers as they should continuously be producing opinions or evaluations that can either be agreed or disagreed with.

However, various techniques can be employed to minimise these effects and facilitate an atmosphere where respondents feel confident in giving their honest opinion, even where it forms a dispreferred response. This can be achieved by:

- Using participants who have similarities - either in their interests or demographically.
- Explicitly stating that disagreement is normal in this context.
- Focusing on disagreements and exploring the reasons behind them.

Although the natural course of conversation tends towards agreement, employing these approaches should ensure that if participants have conflicting opinions they are more likely to be voiced. Where such a disagreement occurs it is possible that the first speaker may downgrade or re-evaluate their initial assessment to try and negotiate a path back to a shared consensus (Pomerantz 1984). This means that hedging or moderation that has been attached to the evaluation may be as a result of the interactional structures at force rather than simply being a reflection of the weakness of their opinion.

The act of disagreement has been studied within the framework of Speech Act Theory (Sornig 1977, 1979), Discourse Analysis (Schiffrin 1994), and Conversational Analysis (Pomerantz, 1984; Sacks 1987; Kotthoff, 1993) but this research uses Politeness Theory as expressed by Brown and Levinson (1987) as its starting point. Sornig (1977: 361) argues that 'the essential characteristic of any expression of disagreement is that it is a reflection of a preceding act which must have been decoded first, and above all, must have been doubted in some of its detail'. While it is of course true that for there to be

something to disagree with, a different opinion or evaluation has to have preceded it, this tells us nothing about the reasons or formulations of a disagreement. So, while early work in Speech Act Theory explored some areas within 'disagreement', it fails to provide an explanation or discussion of the different ways in which any given speech act can be formed. It can be useful in terms of understanding the aims of a communicative act, but it does not necessarily explore or comment on the sociological reasons behind the different manifestations that they may take. Both Discourse and Conversation Analysis also act to describe the linguistic output rather than exploring any social variables and their subsequent influences. Schiffrin's (1984) paper on 'Argument as Sociability' and Kitzinger's (1994) research on refusals incorporate the social and interactional nature of communication but this is more in relation to the context of their research, than the influence it has on the content. Kitzinger (1994: 293) argues that 'Conversation analysis shows that refusals are complex conversational interactions, incorporating delays, prefaces, palliatives, and accounts'. While it is of course interesting to look at the contents and structure of any given communication it does not attempt to provide an account of the social negotiation and interaction that is taking place. In contrast to this, politeness theory, as proposed by Brown and Levinson (1987) does explicitly explore the connection between social influences and relationships and the linguistic output and this is discussed below.

2.1.1.3.2 Power

Fairclough (1993) argues that not enough importance is placed on the power relationships between different speakers and that this can have an effect on their ability to take control of areas such as topic or floor management. This is particularly relevant in an interview situation where the interviewer is inherently more powerful and as such might be expected to have a higher level of control of topic management and turn-taking. Brown and Levinson (1987: 71-83) argue that power differences between speakers determine whether direct or indirect approaches to disagreeing are used. They suggest that less direct strategies of disagreement would be used when there is more social distance between speaker and addressee, when the speaker has less power than the addressee, and when the severity of disagreement is greater. Any disagreements within

the context of my research are not likely to be heavily face threatening as the participants are unlikely to hold extremely strong opinions on the stimuli. Brown and Levinson propose a formula for calculating the 'weight' of a face threatening act as a sum of the social distance of the interlocutors + the relative power of interlocutors + rating of imposition. Although they accept that these ratings are culturally and situationally determined, they argue that the three acting together '...seem to do a remarkably adequate job in predicting politeness assessments' (1987: 80).

In addition to Brown and Levinson's work, the association between language use and power and social distance have been studied from many different perspectives including requests (Blum-Kulka et al., 1985; Lim and Bowers, 1991; Holtgraves and Yang, 1992), apologies (Olshtain, 1989; Holmes 1990) and disagreement (Beebe and Takahashi, 1989). However, while the link seems to be fairly robust, Spencer-Oatey (1996) has called for further investigation into this relationship. She argues that there are two conceptual issues in terms of '...problems of terminology and doubts over the unitary nature of the dimension of 'distance' (Spencer-Oatey (1996: 2). She points out that the following tags have all been used as equivalent to 'social distance'; 'solidarity', 'closeness', 'familiarity' and 'relational intimacy'. 'Power' has been referred to as: 'social power', 'status', 'dominance' and 'authority'. This raises the question of whether researchers conceptualise social distance and power in the same way even when using slightly different terms. For example, distance, closeness and familiarity could potentially refer to one or more of the following: frequency of contact, length of acquaintance, amount of self-disclosure, and amount and type of affect. Spencer-Oatey suggests that these different interpretations of the concept of 'social distance' are likely to have an effect on the way in which it is deemed to have a causal relationship with language choices. There is little argument that it is important for terminology to be used correctly and consistently. However the sheer volume of empirical research that Spencer-Oatey herself refers to, suggests that while there might be a slight lack of clarity or preciseness, the underlying principle that social distance and power differences between interlocutors effects language choices, remains strong. Such loose terminology is indicative of newly emerging and developing fields and is not necessarily problematic.

In settings with inherent power differences, such as moderator / participant interviews, the less powerful participant is likely to use more hedges and mitigation (Fairclough 1989). Although it is not explicitly stated, the implication is that in terms of face, at least, the cost of disagreeing with a more powerful interlocutor is greater than disagreeing with a speaker of less or equal power.

While it was made clear to the participants that the moderator has no affiliation with the stimuli, it is likely that some association remained, in turn affecting the participants' responses. That the interviews took place at Unilever's offices may also have contributed to this. It is therefore possible that the participants felt that the preferred response would be to like the stimuli with a disliking being a dispreferred response. This means that questions such as '*Ok so do what you think of this one*' or '*Have a look at this one next*' might not be interpreted by the respondents as referential. Instead, due to the context they might perceive them as something more akin to '*Do you like this one?*' This would mean that there would be a preferred response which would be to 'agree' with the question and confirm that the stimulus was liked. Negative evaluations contained within a seemingly neutral question from the moderator could therefore perhaps be considered to have a higher cost than a positive evaluation to a moderator question.

In addition, where a moderator asks a question without specifically selecting the next speaker, the respondents have the opportunity to self select. As stated above, where a speaker is directly selected by the current speaker, there is a lower level of risk which is therefore indicative of a lower level of evaluation. Where the respondent self selects from an open moderator question there is a greater level of risk, particularly when giving a negative evaluation that may be perceived by the speaker as a dispreferred response.

Rees-Miller's (2000) study of disagreements between students and academics also explored the way in which power differences affected the construction of disagreements in conversation. Her findings ran counter to her expectations and showed that students

softened their disagreements with professors only marginally less than when disagreeing with their peers. However she then goes on to explain that this is perhaps because in this specific context, disagreement is seen as a positive thing as it shows that the students are engaging with the topic and getting involved. It could perhaps be argued that this has some similarity with the interview setting of my research though there are likely to be fewer instances where it occurs. In terms of cost, Rees-Miller (2000: 1095) also points out that for her research at least, 'students who disagree must also assess the possibility of incurring retribution when the time comes for the professor to assign grades for the course'.

Where a respondent takes control of the floor management against an inherently more powerful speaker, it may be possible to make inferences about the inherent strength of any evaluative content included in these turns. Particularly when giving a dispreferred response to an utterance made by the more powerful speaker as this would be perceived as having a high cost. In contrast to the difficulty of providing a dispreferred disagreement response, preferred agreement responses carry far less potential cost to the speaker. There is an expectation from the initial speaker that a preferred response format will be forthcoming and agreeing with the initial assessment is far less threatening. Because of this, a preferred response has a lower cost to the speaker than a dispreferred response and is therefore usually immediate and unmarked. This is in contrast to a response to be taken as disagreeing which will typically be delayed, prefaced, or modified (Sidnell 2010: 78). It follows, therefore, that if an initial suggestion is met with silence or an extended pause, this may be an indication of an as yet unstated dispreferred response (Pomerantz 1984). Indeed, Cameron (2007: 95) suggests that because the pattern is so consistent, and the alternative so distinct, disagreements and refusals are interpreted as soon as an initial hesitation is registered.

Where a participant disagrees with a previous assessment but wants to lower the strength of their dispreferred response, there are several options open to them. As noted above, the turn may be delayed, prefaced, or modified. Pomerantz (1984: 58) suggests that the second speaker can also claim insufficient knowledge to be able to agree, before going on

to disagree. 'One of the ways of warranting a declination, then, is to deny the proper basis, that is, sufficient knowledge, for its production.' This can be seen in the following example from the data:

22. Speaker A: *I think that would kind of work if you lived in one of them really swanky studio type flats where you and if you just wanted nice things on display to sort of show off.*

Speaker B: *Well I wouldn't know about that but I don't think I'd want I wouldn't have that in my bathroom.*

Here, the second speaker uses this technique to downplay their disagreement. The '*Well I wouldn't know about that*' refers to the previous assessment that it would '*kind of work if you lived in one of them really swanky studio type flats*' and acts to mitigate the following statement that they would not want the pack and do not like it. Where a dispreferred response is given, it is also possible that a weak preferred response precedes it. This may be provided to ensure that the speaker gets hold of the floor and once this has been achieved they will then give their real opinion of the previous speaker's evaluation. Using this weaker agreement also acts to mitigate and delay the ensuing dispreferred response. In the following example from the transcripts, speaker A states that they dislike the whole concept of the pack due to one particular feature. Speaker B starts their turn by agreeing *Yea I see what you mean* before following it up with a more positive evaluation about a different aspect of the design - *I quite like the style of the erm the grips on that.*

23. Speaker A: *Oh I really don't no I don't like that one at all it's far too bulky.*

Speaker B: *Yea I see what you mean but I was going to say that I quite like the style of the erm the grips on that.*

Example 23 shows how a disagreement can be prefaced by a weak agreement that acts to allow the speaker to get hold of the floor, mitigate their disagreement and also delay it.

Sornig (1977: 362) argues that 'because disagreement is based upon and takes its start from what has been said before I assume disagreement to be easier than asking for new information or even [providing] certain kinds of answers'. However, this fails to take into account Brown and Levinson's (1987) notion of Face discussed below and I would therefore argue that although a disagreement is indeed rooted in what has preceded it, any potential friction to the social and interactional structure of the communication far outweighs this 'ease'.

For agreement responses Pomerantz (1984) suggests that they can be split into three further categories and can constitute an upgrade, a downgrade, or a 'same' response. An upgrade agrees with the initial assessment but uses a stronger evaluative term where graded sets of descriptors are used. The following is an example of an upgraded agreement containing a stronger evaluative term:

24. Speaker A: *Yea that's nice that.*

Speaker B: *Yea that's great.*

This upgraded evaluation shows a higher level of agreement and can therefore be considered a stronger evaluation than a 'same' or 'downgraded' evaluation. The following is an example of a downgraded agreement realized through a weaker evaluative term:

25. Speaker A: *And the idea of those grips is fantastic really you know for when you're in the shower or wherever and you get a bit slippery.*

Speaker B: *I don't think I wouldn't buy something because of that but they are good yea.*

The final factor I consider when exploring encoded evaluative strength is whether or not the turn constitutes what Gardner (2001) terms 'response/non-response tokens' and this along with an example of a 'same' agreement is discussed below.

2.1.1.4 Content and Non Content turns

In his book on listener responses Gardner (2001: 1) argues that across early work in linguistics the focus has primarily been on the producer rather than the receiver. Though he acknowledges that this started to change through the work of CA he suggests that more can be done to unpick what has broadly been termed 'backchanneling' (Yngve 1970: Duncan and Fiske 1977). These include items such as 'yea', 'uhuh', 'mm', 'right', 'really', 'great', and so on. Gardner (2001) depicts a typology of these markers including continuers which act to note agreement with the speaker and give them back the primary role in the conversation, acknowledgements which work to express agreement or understanding between a speaker and a listener, and repairs that ask the speaker to rephrase or repeat an idea or question. However it is those which he classifies as 'acknowledgements' which are of most interest to this study. Gardner (2001: 2) describes acknowledgements as those which 'claim agreement or understanding of the previous turn (e.g.) Mm, yea'. Given the structure of the research interviews, outlined in detail in section 5.2, the majority of these responses are likely to be between respondents rather than between the moderator and a respondent. However at various stages it is likely that the moderator will summarise what has been said previously to continue a line of questioning, to clarify a point, or to elicit a response from a different participant on the same subject. Myers (1998) does suggest that summaries should be avoided as they might be leading but they can be utilised as a useful linking tool. Under Sinclair and Coulthard's model they would be considered a follow up and Sinclair (1994: 86) states that they are '...distinctive, and restricted to areas of discourse where one participant has the right to evaluate the behaviour of another.' Follow-up moves can therefore be seen as a mechanism of coherence within the discourse. It is in these situations where a respondent to moderator acknowledgement might take place.

Despite the attention and progress within CA, Gardner (2001: 20) argues that a consensus is lacking as to the meaning of these kinds of utterances. He lists nine questions that he suggests research needs to address but admits that 'some of these questions would be very difficult to answer satisfactorily' as 'each has to be interpreted according to its

placement within a sequence of talk'. Finally he suggests that 'this will involve an interpretation according to intonation contour and other prosodic features for example duration, pitch height and amplitude'. Goodwin (1986: 210) also states that 'their impact appears to hang as much as anything else on their intonation contour and prosodic form' and Muller (1996: 136) argues that 'prosody can change the impact of even the most minimal or 'neutral' of responses'.

For reasons discussed in section 2.4 a prosodic analysis is not undertaken in this research making it difficult to untangle the different meanings contained within these kinds of utterances. I therefore take a broad approach in keeping with Gardner's (2001: 3) assessment that these utterances provide 'evidence of the stance that the recipient in the talk is taking at that moment' and Gerhardt and Beyerle (1997: 384) observation '...that Mm-hmm and Uh-huh have positive valence [and] yeah and yes show affirmation and agreement'. For these reasons I argue that where acknowledgement such as 'yea', 'uhuh' and 'mm' are made, they are indicative of a very low level of agreement with the prior evaluation. Conversely a simple 'no' utterance is viewed as being indicative of a low level disagreement. Due to the discursive nature of the research data these non content disagreement turns are unlikely. Indeed, the only occurrence of a participant making a non-content disagreement turn was to the moderator where the participant was using humour and sarcasm to try and downplay their dispreferred response.

26. Moderator: *So you think that it's a bit cheap looking?*

Speaker B: *No.*

[laughter]

Moderator: *Earlier before when you said.*

Speaker B: *Well I said it was maybe a bit you know cheapy looking.*

In their next turn the participant accepts that they did call it cheap looking but attempts to reduce the strength of their evaluation with the use of weakening markers such as 'maybe' and a 'bit'. Apart from this usage it is unlikely that that a participant would give

a dispreferred response to a speaker with more power in this market research interview setting.

A distinction is though made between acknowledgements produced in response to the more powerful moderator than to a more equal co-respondent. Agreeing with a speaker who has a more powerful role in the interaction is seen as having a lower level of inherent strength than agreeing with an equally powerful speaker.

I also argue that these agreement tokens, lacking in additional information, can be compared to Pomerantz's idea of a 'same' evaluation as it neither up nor downgrades the evaluation. I therefore suggest that a non-content agreement turn should be considered as being indicative of a lower level of inherent strength when considered as evaluative talk. Indeed, Pomerantz (1984: 69) suggests that downgrade or same agreements frequently precede a disagreement sequence and therefore a '...same evaluation agreement may be considered a kind of weak agreement.'

27. Speaker A: *We're used to more sort of thinner canisters.*

Speaker B: *Yea.*

In this section I have argued that there are four factors that should be analysed to better understand the inherent strength attached to evaluative language. Firstly I introduced the concept of turn taking, arguing that an initiation turn is indicative of a stronger opinion than a response turn. Within that I suggest that a self – selected turn is also indicative of a strong opinion. Particularly in relation to the second factor of preferred or dispreferred responses and the perceived context of having to provide a preferred response to the moderator. The third area to be considered is the relative power differences between the speakers. I argue that responding to the moderator is more threatening and therefore indicative of a stronger opinion than responding to a fellow research participant. The final factor is that of whether or not the turn is considered a content or non-content turn. I argue that content turns are indicative of a stronger evaluative force than non-content turns.

One further linguistic feature also has the potential to provide further insights into a speaker's strength of feeling or conviction with regards to what they are saying. The use of tag questions has received much attention and the extensive literature on tag questions provides much information on their use from both a semantic and pragmatic perspective. In the analysis of the data of this research, tag questions are used as an auxiliary method of exploring the strength of an evaluation. The following section will provide an overview of the literature with regards to tag questions and the roles that they can play within communications.

2.1.1.5 Tag Questions

There are several different models of the different types of tag questions but the area I am most interested in is the notion that they can be used to express uncertainty or lack of engagement with the preceding evaluation. Holmes (1995: 80) suggests that they can 'express genuine speaker uncertainty rather than politeness' and Tottie and Hoffmann (2006: 297) state that they can be regarded as 'hedges on the preceding proposition'. The concept and role of hedging is discussed in more detail in section 2.2.3 but relates to the way in which a speaker can modulate their explicit perception of certainty and accuracy and also manage interpersonal relationships. However not all tag questions indicate doubt in the mind of the speaker. Algeo (1990: 445) suggests that 'a more frequent use of tag questions is not to seek information but to draw the person addressed into the conversation'. Holmes (1995: 81) proposes two major categories of tag questions; Epistemic Modal and Affective with the latter split into three sub-categories of Facilitative, Softening and Challenging:

- Facilitative tags '... are examples of hedges which serve as positive politeness devices. They invite the addressee to contribute to the discourse'.
- Softening tags are described as 'negative politeness devices, used to attenuate the force of negatively affective utterances, such as directives...and criticisms'.

- Challenging tags are ‘confrontational strategies which may pressure a reluctant addressee to reply or aggressively boost the force of a negative speech act’.

One of the main ways in which the classification of a tag question can be determined is through the intonation patterns that are used by the speaker. Tottie and Hoffmann (2006:299) suggest that the ‘multifunctionality of tags is a problem for any classification’ and this is also supported by research by Holmes (1983), Cameron et al (1989) and Coates (1996). Coates (1996: 196) explains it thus:

Two tag questions [can] use the same words, but express very different meanings. The first is a typical confirming-the-shared-world tag which expects no reply; it has falling intonation. The second is an information-seeking tag with rising intonation, which does seek a response’.

Looking at the use and distribution of tag questions might facilitate a further level of analysis of the strength of a speaker’s evaluation and could provide a layer of analysis in their own right as discussed in section 5.4.

As discussed throughout this section, a further factor impacting on both the construction of communication and the encoded level of evaluative force is that of Politeness Theory and this is discussed in greater detail below.

2.1.2 Politeness Theory

Brown and Levinson’s approach (1987) to the domain of Politeness Theory has become an influential area within applied linguistics and more specifically pragmatics. They frame their account of politeness resources within a theory of politeness or face that develops Grice’s theories of conversational maxims (Grice 1975). Grice suggested that

all speakers operate within the confines of a 'co-operation principle' that is comprised of four parts.

Maxims of quantity

- Make your contribution as informative as required.
- Do not make your contribution more informative than is required.

Maxims of quality

- Do not say what you believe to be false.
- Do not say that for which you lack adequate evidence.

Maxim of relation

- Be relevant.

Maxims of manner

- Avoid obscurity of expression.
- Avoid ambiguity.
- Be brief.
- Be orderly.

Brown and Levinson explain apparent departures from these maxims in terms of their notion of politeness. This is of relevance to this research as many of these departures are realised through the use of lexical items which are introduced in the discussion of stance, evidentiality and epistemic modality below. For example, expressions of doubt; *perhaps*, *I guess*, *I think*, expressions of certainty; *definitely*, *for sure*, *must*, expressions of deduction; *it seems to me* and quotatives / hearsay; *it's said* can all be used in contravention of the maxims of quality and manner. Brown and Levinson use a broader

category than is typical under other approaches to evidentiality or epistemic modality. They also include values which Lakoff (1972) termed 'hedges' which they define as a 'particle, word or phrase that modifies the degree of membership of a predicate or noun phrase in a set; it says of that membership that it is partial, or true only in certain respects, or that it is more true and complete than perhaps might be expected' (Brown and Levinson 1987: 145). Such items include things that as discussed above, have elsewhere been included as examples of stance or evidentiality; '*it's a sort of spaceship*', '*Yea I'm fairly sure I wouldn't buy that*', '*well it's not a real problem*'. They also include items which Labov (1984) termed intensifiers or amplifiers; '*No no I do not like that at all*', '*yea that's really good*', '*That is just you can't it's completely impractical*'. It is interesting to note that the broader domain that is used by Brown and Levinson, resembles much of what Martin and White (2005) include under the domain of ENGAGEMENT as introduced in section 3.2.2.

2.1.2.1 Face

Brown and Levinson base their argument on the proposition that certain human wants or desires are fundamental to social interactions. They make the assumption that all competent communicators in a society have, and know each other to have both positive and negative face and that these should be maintained during communication. They derive the notion of face from Goffman (1967) describing it as something that 'can be lost, maintained, or enhanced, and must be constantly attended to in interaction' (Brown and Levinson 1987: 61).

- Positive face is concerned with an individual's self esteem and the sense that they are approved of and that people share and endorse their values.
- Negative face is concerned with an individual's desire 'to be unimpeded', 'the basic claim to territories, personal preserves, rights to non-distraction - i.e. freedom of action and freedom from imposition' (Brown and Levinson 1987: 61).

These face needs can also be understood in terms of Maslow's hierarchy discussed in 1.1.1., with human needs being met and managed through communication. Indeed, Yngve and Wasik (2006: 270) argue that a '...human need may well motivate and influence a particular instance of human communicative behaviour.' Certain communications are understood to put one or both of these dimensions of face at risk, that is, they constitute face threatening actions. In order to minimise the social damage put at risk by such threats, speakers resort to communicative strategies of 'face saving'. It is under this pressure for 'face saving' that Grice's maxims are sometimes put at risk. Therefore, on the face of it, some communications are not sincere, or maximally efficient, relevant or mannered. These departures, such as hedges, are understood as being motivated by maintaining face and politeness. Therefore the hearer is likely to understand that the apparent indirection and redundancy in a request acts to signal the speaker's concern for the hearer's negative face, as a way for the speaker to show that they wish to minimise the imposition brought about by their request. Hedging and its role in negotiating both positive and negative face is discussed in greater detail below in section 2.2.3.

Section 2.1.2 has looked at some of the different effects that maintaining politeness and face can have on a speaker's utterances. In section 2.1.3 this is then combined with the discussion on turn taking, preference and power, and content and non-content turns to propose a framework for exploring the inherent strength of evaluations based on both the structural and social elements that are involved in the production of any given utterance.

2.1.3 Proposing a Scale of Inherent Strength Based on Structural and Social Factors

Throughout this section I have argued that both the social and structural elements of communication can be seen to determine the inherent strength or weakness that is encoded within an evaluation. I propose that more importance can be placed on some

evaluations than others, based on the difficulty or cost associated with making that evaluation. This section now proposes a framework for exploring the inherent strength of evaluations based on the following factors:

- Where the turn fits in the turn-taking structure
- Whether it a preferred or dispreferred response
- Whether it is a response to the moderator or a co-respondent
- Whether it is a content or non-content turn

Examples of each of the different turn-types are shown and where there were no occurrences within my data set a potential manifestation is shown.

1. Unsolicited Evaluation: Initiation evaluation (I)

Speaker B: *I think that's quite bulky.*

2. Response to Moderator: Non-Content Disagreement Turn (RMNCD)

Moderator: *So you think that it's a bit cheap looking?**

Speaker B: *No.*

*No examples in transcripts – potential realisation given.

3. Response to Moderator: Content Disagreement Turn

Moderator: *This one's too big?*

Speaker A: *No actually I think it's the right sort of size.**

*No examples in transcripts – potential realisation given.

4. Response to Participant: Non-content Disagreement Turn (RPNCD)

Speaker A: *That's better don't you think?**

Speaker B: *No.*

*No examples in transcripts – potential realisation given.

5. Response to Participant: Content Disagreement Turn (RPCD)

Speaker A: *Erm I guess the opposite to fresh or cool something that's more just just just plain if you like nothing that you'd smell and think wow god I like that or what's that like.*

Speaker B: *No I think it would be cool.*

6. Response to Moderator Question: Content Turn (RMC)

Moderator: *Ok er in what way is it too bulky?*

Speaker B: *It's just too big.*

7. Response to Participant: Content Agreement Turn (RPCA)

Speaker A: *The handgrips I like the handgrips.*

Speaker B: *Yea mmm I think that'd be quite handy for a bathroom if it it things can be a bit slippery.*

8. Response to Participant: Non-content Agreement Turn (RPNCA)

Speaker A: *We're used to more sort of thinner canisters.*

Speaker B: *Yea.*

9. Response to Moderator: Content Agreement Turn (RMCA)

Moderator: *So I mean would that be more expensive then?*

Speaker A: *Yea I'd say more expensive yea.*

10. Response to Moderator: Non-content Agreement Turn (RMNCA)

Moderator: *So you think that that would be cheaper?*

Speaker B: *Yea.*

This scale is used to form part of the tagging of the data as detailed in section 5.4.2.

Having looked at different areas of discourse that have been explored to understand evaluative talk, the following section will focus on grammar and evaluation exploring Stance, Modality, Hedging and Evidentiality.

2.2 Grammar

Perhaps the first observation to make within this section is that much previous research into grammar and evaluation has focused on specific areas that fall under the auspices of Stance, Modality, Hedging and Evidentiality. Labov's (1984) paper on 'intensity' and Chafe's (1982) research on 'emphatics' exemplify the trend for sticking to the analysis of individual markers. Intensity is defined as the 'emotional expression of social orientation toward the linguistic proposition: the commitment of the self to the proposition' (Labov 1984: 43-44) and emphatics simply '...mark the presence of certainty towards a proposition' and thus differ from 'amplifiers' which 'indicate the degree of certainty' (Biber & Finegan 1989: 94). As a result there is great overlap regarding the terminology and concepts that different research uses. Hunston and Thompson (2003: 2-3) suggest that 'different contributors use different terms: Martin talks of *Appraisal*, while Conrad and Biber talk of *stance*' and that where '...the traditional term for this is *modality*... there is a good deal of overlap with the concept of

evidentiality'. This section attempts to untangle this to some extent using examples from the research data.

Thompson and Hunston (2003) make a distinction between two kinds of evaluation. Firstly, a speaker may be valuing or taking a position for or against a proposition or an idea:

29. *Yea I like the idea of a sort of you can just put it down the side of your travel bag or whatever.*
30. *Yea I'd say women are more likely to want this sort of smaller thing.*

This could be a response to a statement such as *'People well women just want smaller deodorants'* or a question such as *'What do you think of smaller deodorants?'*

The second kind of evaluation is in relation to an object or entity where the opinion is given in response to exposure to the stimulus:

31. *Yea I quite I do quite like that.*
32. *I just think it'd be a bit of a talking point I think.*

This might suggest that all evaluation is simply a response to something else. Indeed Labov (1972) suggests that evaluation can be defined by the comparison or contrast of something to a perceived norm. However, this does not necessarily mean that it has to be a response to an explicit comment or question. Instead, the context or setting can be viewed as initiating the communication, particularly in a context such as a market research or new product development interview. In such a setting it should be clear that the participant is being invited to give their view, even if there is not a specific proposition or statement from the moderator. Therefore, even if an utterance is viewed as being an initiation from a turn-taking or conversation analysis perspective, it can still be interpreted as a response.

For Fairclough (2003: 164) the distinction between entities and propositions is the speaker's commitment to 'what is true and what is necessary' and to 'what is desirable or undesirable, good or bad'. However this distinction becomes less clear when the proposition or idea is directly related to the object or entity. If one participant makes the proposition that *'this one it's this one is definitely way too big for your hand it's useless'* and a second speaker says *'yea it is yea kind of it would be better if it were smaller'*, then this is evaluating both the proposition and the object or entity about which that proposition was made. The second speaker's use of *'yea it is yea'* shows that they are agreeing with the first speaker. For Fairclough then, this is an assessment of a proposition and its truthfulness. However this is followed by *'it would be better if it were smaller'* which is an assessment of the object and therefore of what is (un)desirable. Thompson and Hunston (2003: 3-4) also recognize the potential for these formulations stating that 'these differences are to some extent misleading, in that both types of expressions of opinion do share certain structural possibilities [and] this suggests that there is a fair degree of common ground'.

Thompson and Hunston (2003) also suggest that this distinction between objects and entities and ideas and propositions affects the way in which the evaluations are grammatically realised. Entities are more likely to be evaluated through the use of adjectives whereas propositions are more likely to make use of categories such as modal verbs. These two approaches, to differentiating types of evaluation, can be understood in terms a 'separating' or a 'combining' approach. The separating approach concentrates on the differences, gives each type a separate label and in the main analyses them as separate phenomena. The Hallidayan approach can be seen as representative of this method which also includes Martin (1997; 2002) and Eggins and Slade (2001), are also amongst its proponents. The alternative, combining, approach is taken by Hunston and Thompson (2003) as well as Conrad and Biber (2000) and Biber and Finegan (1989). Although this research is concerned with the evaluation of entities and objects (the stimuli), I argue that the distinction is not necessarily clear cut. There are likely to be situations where one participant is responding to a proposition or idea made by either the moderator or the other participant. Therefore both evaluative types will be explored. The following four

sections on Stance, Modality, Hedging and Evidentiality introduce each individually, in keeping with the separating approach outlined above. At the same time, however, I acknowledge that there are significant levels of overlap between these areas, particularly at the higher levels which are mostly discussed here.

2.2.1 Stance

Baratta (2009: 14007) argues that Stance ‘would appear to fall along a continuum, ranging from hedging, in order to reveal doubt about one’s personal claims, to offering a personal opinion regarding the claims of others.’ Again this highlights the level of overlap that exists when trying to unpick these areas. Conrad and Biber (2000: 58) are concerned with grammatical expressions of stance. They define these as ‘the use of a grammatical device to provide a personal framing for some other proposition’. Stance is used to cover the expression of feelings and assessments (evaluation) across three domains:

- Epistemic stance, passing judgement on the certainty, reliability or limitations of a proposition
- Attitudinal stance, conveying the speaker’s attitudes, feelings or value judgements
- Style stance, passing judgement on the manner in which an evaluation is made.

Berman (2005: 107) focuses on similar areas but uses the terms *Epistemic*, *Affective* and *Deontic Stance* respectively. Conrad and Biber (2000: 73) use a corpus approach to study the different ways in which speakers use adverbials to mark their stance across different genres and registers. Perhaps the most relevant finding for this thesis is their (2000: 73) claim that:

‘Conversation was found to have a high frequency of stance adverbials marking doubt, imprecision and actuality.... it is clear that they have

important social functions beyond simply marking the speaker's stance.'

This is a view shared by Martin and White (2005) who also propose that such markers play an important role in negotiating interpersonal relationships and should not therefore be interpreted solely as indicative of certainty. Instead their role in the construction of such interpersonal relationships should also be considered as this can also be used to infer evaluative strength. This is in keeping with their view on Hedges which they also view as having interpersonal significance and this is discussed further below. Conrad and Biber (2000) also suggest that the use of adverbial markers of stance is markedly different depending on the communicative purpose of an utterance.

Despite concentrating on these grammatical realisations, they agree that evaluation can also be realised through individual lexical items. These can take the form of verbs '*yea I like that*', adjectives '*it's too it's too thick*' and nouns '*it's a bathroom shelf object*'. However, they restrict their study to adverbials as they are interested in grammatical devices used to frame a proposition and therefore are not seeking to explore value laden lexical choices.

Epistemic stance relates to the speaker's own certainty about what they are saying and several different sub-categories can be distinguished. These relate to the certainty of the proposition:

33. So that would be *perhaps* quite useful.
34. Yea that would *probably* make me think of that.
35. Erm what would I say it was for *definitely* a younger age.

The reality of the proposition:

36. It *actually* leads you to pick it up.
37. I think *the fact* that it's unusual would go for the younger audience.

38. I do not *really* think I'd have a feel of it in the store.

The preciseness of the proposition:

39. Cos it is quite *sort of* funky funky design.

40. You'd expect it to be a *kind of* invigorating refreshing smell.

The source of the information:

41. Though *apparently* smaller ones are more erm popular.

42. Well it's *evidently* not been thought about very much.

The limitations of the proposition or perspective in which it is relevant:

43. Well *from my point of view* I'd definitely want it to stand up.

44. You're usually in a hurry aren't you *most of the time* you just want to pick it up and use it.

Attitudinal stance adverbials cover a wide range of meanings expressing attitudes, feelings and expectations but it is less obvious to see how they might fit into discrete categories. Examples include the use of words such as *fortunately, surprisingly, sensibly, amazingly* and so on.

As there are no examples of this in my data, the following four invented examples show how style stance indicates the way in which the speaker is presenting their information.

- *Briefly*, I would like to introduce this new scheme.
- *More simply put*, one is the cause of the other.
- This is *literally* going to kill him.
- I'm fine, *honestly*.

This can act to both state the way in which they are presenting their ideas and also the way in which they intend them to be interpreted. Although Biber and Conrad's (2002) notion of style stance only relates to the framing of an individual's own propositions, it is also possible to see examples of a speaker commenting on the style stance of their interlocutor.

45. Speaker A: *This one it's this one is definitely way too big for your hand it's useless.*

The second speaker responds with

46. Speaker B: *Well I wouldn't put it like that but yea it is a bit big.*

In response to an initial statement, the second speaker provides an evaluation of the previous speaker's proposition, the way in which they made that proposition and also the object about which the proposition was made. Other examples of style stance include the following phrases:

47. *I don't think it's as simple as that.*

48. *Yes I mean that's the way I was trying that's how I wanted to say it'.*

Biber and Finegan (1989: 94) use the term 'stance' to encompass what they call the exploration of 'the lexical and grammatical encoding of both evidentiality and affect in English'. One of the main aims of their research was to 'survey previous research and individual texts in order to identify potentially important grammatical markers of stance' (Biber and Finegan 1989: 98). They found six affect groups; adjectives, adverbs and verbs, either positive or negative.

	Adjectives	Verbs	Adverbs
Positive	I feel fortunate; it is amazing that	I enjoy; it pleases me	Happily, luckily
Negative	I am shocked; it seems unnatural	I dread; it scares me	Alarmingly, sadly

Table 2 Six different affect groups (Biber and Finegan 1989)

They then made a statistical analysis of the frequency of occurrence of these groups across different genres. Their main findings were that nearly two thirds of the texts were unmarked for stance, and that the remainder were ‘organized by both semantic and grammatical criteria’. The suggestion that only a third of texts are marked for stance is at odds with subsequent research that suggests that evaluation or stance is far more pervasive than this (Barton 1993; Stubbs 1996; Hunston and Thompson 2003). Martin and White (2005:92) take this more pervasive view even further stating that their approach locates them in a tradition in which ‘...all utterances are seen as in some way stanced or attitudinal’. This suggests that much evaluation falls outside the grammatical markers of stance that were the main point of study for Biber and Finegan’s (1989) research and adds weight to the argument that a full and proper analysis of evaluation needs to encompass as many different potential markers as possible. Precht (2003a) argues that markers of stance are indicative of an ingrained system or shorthand for expressing emotions and attitudes. She suggests that the use of stance markers is culturally specific and that although there are more than 1400 different stance markers available to speakers; we use less than 10% of these for more than ninety percent of our stance expressions (Precht 2003b). Each of Epistemic, Attitudinal and Style Stance are areas that are used in the analytic approach of this thesis. While the terminology and exact definition that is used differs as a result of the main framework of analysis that is discussed below in Chapter 3, accounting for the way in which speakers vary the extent to which they show their own certainty about what they are saying, express attitudes and feelings and comment on other speakers’ observations are central to the overall analytical approach that is taken. Although this section has attempted to explore previous ways in which stance has been discussed in the literature, Hunston and Thompson (2003:2) perhaps best sum it up declaring that ‘stance is not always a straightforward matter’.

2.2.2 Modality

Halliday (1994: 88) describes Modality in relation to Polarity where Polarity is ‘the choice between positive and negative, as in *is/ isn’t, do/ don’t*.’ He then proceeds to point out that:

'The possibilities are not limited to a choice between yes and no. There are intermediate degrees: various kinds of indeterminacy that fall between, like 'sometimes' or 'maybe', These intermediate degrees, between the positive and negative poles, are known as MODALITY'.

Halliday also makes a distinction between what he terms Propositions and Proposals. A Proposition is related to the assertion or denial of an utterance with positive and negative manifestations such as:

- 29. *That's too big.*
- 30. *You couldn't even stand that up.*

Propositions come with intermediate possibilities relating to degrees of probability such as 'probably', 'possibly' or 'certainly' and degrees of usuality such as 'sometimes', 'usually' and 'always'. A Proposal is related to the carrying out of an action, what Halliday (1994) terms 'goods and services'. Here the positive and negative realizations are related to doing or not doing the action:

- 31. *I'd buy that.*
- 32. *I wouldn't even pick that up.*

For a Proposal the intermediate possibilities vary according to whether it is a Command or an Offer. For a Command they describe the level of obligation: 'may', 'should', or 'must' and for an Offer they relate to levels of inclination: 'might do', 'could do' or 'will do'. Fairclough (2003: 165) summarises this stating that 'the question of modality can be seen as the question of what people commit themselves to when they make Statements, ask Questions, make Demands or Offers'. Verschueren (1999: 179) describes Modality as '...the many ways in which attitudes can be expressed towards the pure reference and prediction content of an utterance'. There is then, clear overlap between Stance and

Modality in the way that Hodge and Kress (1988) use the term 'stance' for the same purpose, that is, acting to show the degree of affinity the speaker has towards their utterance.

Fairclough (2003) proposes that the idea of modality should go beyond simply encompassing modal verbs and must instead include all tools through which a speaker modulates their attachment, or detachment, to a proposition. He highlights the work of Hodge and Kress (1988) as an example of this extension and supports their inclusion of verbs of appearance such as 'seem' or 'appear', in addition to the adverbs that are more traditionally used as markers of modality. Fairclough (2003: 171) also proposes the inclusion of hedges such as 'sort of' and 'kind of' in addition to some of the prosodic features that will be discussed below in section 2.4.1, arguing that 'intonation and other aspects of oral delivery are also relevant to a speaker's degree of commitment'.

While both Proposals and Propositions were evident in the research data, the more hypothetical nature of the questioning meant that Propositions were more prevalent. In relation to Propositions White (2003) builds on Martin's (1992, 1997) notion of Engagement and Bakhtin's (1981, 1986) and Volosinov's (1973) work on verbal communication and takes a dialogic approach to the understanding of the evaluation of propositions or ideas. This notion of Engagement is a catch-all term for resources of intersubjective positioning and is introduced and discussed further in the introduction to Appraisal Theory in section 3.3. Bakhtin's influential notion of dialogism proposes that all language is produced against the backdrop of what has been said before and in anticipation of what might be said in the future. This is relevant here as this dialogic approach places White at odds with much of the prior research in this field. White (2003) cites work by Chafe and Nichols (1986), in addition to work on hedging by Markkanen and Schroder (1997) as being indicative of the non-dialogic view point. These studies, and others in the modality and evidentiality literature, often assume that the:

'...sole function of these wordings is to reveal the speaker/writer's state of mind or knowledge, to indicate that the speaker/writer is uncertain

or tentative and is not committed to the truth value of the proposition’.

(White 2003: 261)

In other words, there is no concern given for external influence and such markers only show the speaker’s commitment to different aspects of an evaluation. This is not, however, to say that all work in this field makes this assumption and Myer’s (1989) study of politeness in scientific articles is one such example that does not. Myers suggests that when criticizing other researchers’ work, this is inherently what Brown and Levinson (1987) term a Face Threatening Act. The use of ‘various hedges to modify statements that could be FTAs’ (Myers 1989: 30) can be considered, he argues, not only an assessment of the work that they are critiquing, but also an attempt to downplay these potential Face Threatening Acts. Having introduced the concept of Modality, the following section continues the discussion on Hedging.

2.2.3 Hedging

Hood (2004) argues that one of the most influential research areas within explorations of evaluative language is that carried out under the term Hedging. (Hyland 1994, 1998, 2000, Myers 1989, Salager-Meyer 1994). Hyland (1998:1) describes Hedging as:

‘Any linguistic means used to indicate either a) a lack of complete commitment to the truth value of an accompanying proposition, or b) a desire not to express that commitment categorically’.

Lakoff (1972: 195) simply defines Hedging as a collection of resources for ‘making things fuzzier’. In common with Stance, a split between epistemic and interpersonal functions are once again highlighted with the former encoding the speaker’s perceptions of their evaluations in relation to accuracy and certainty and the latter relating to the constraints and expectations imposed by actual and imagined interlocutors. Brown and Levinson (1987: 117) view them as interpersonal tools arguing that they are ‘used to

soften FTAs [face threatening acts] of suggesting or criticizing or complaining, by blurring the speaker's intent'.

Whereas the focus on hedges, in the broadest term, has been on their function in relation to an utterance's truth value or epistemic reliability, Brown and Levinson move their usage and understanding of these hedges into the interactional picture. This places them alongside analysts such as Labov (1966, 1972) and Cortazzi and Jin (2000) who also view evaluation as being a primarily social tool. They propose that hedges can be socially motivated as a way of negotiating and maintaining relationships, rather than simply being a result or display of doubt or mitigation. Although Brown and Levinson (1987) suggest that hedging is primarily used with regards to negative face to minimise the imposition, they also propose that in some instances hedging applies in the context of threats to positive face. It can be used to compensate in situations where it might be apparent that one speaker does not share the same values as the other. This is particularly relevant to a paired-depth interview on opinions on topics such as pack design and usability as it would be unlikely that two participants would agree on everything that the other has said. This can be seen in the following example where the second speaker disagrees with the values of the first and uses a hedge to downplay this positive face threatening action.

49. Speaker A: *I think that that would be for older people you know like your Bruts or your Old Spice.*

Speaker B: *Oh no I just to me that would be sort of young and fresh yea.*

White (1998) agrees, suggesting that 'here hedging applies in the context of opinions which may put agreement at risk'. However, where Brown and Levinson argue that the use of hedging maintains the sense that there are shared values even in the face of apparent disagreement, White suggests that this is too simple. Brown and Levinson suggest that the use of a hedge makes the opinion too vague to be face-threatening. While White (1998: 17) acknowledges that 'such values may be mobilised so as to negotiate areas of disagreement' he argues that the underlying, differing, opinion remains the same and any notion of vagueness counter-balancing this is relatively weak in comparison.

Although this distinction can be made at a theoretical level, any attempt to encode and analyse the dual function of epistemic and interpersonal formulations involves some degree of interpretation from the researcher. Hyland acknowledges that ‘in actual use the epistemic and the affective functions of hedges are often conveyed simultaneously and that this indeterminacy prevents the formation of discrete descriptive categories’ (Hyland 1998: ix). Dascal (2001: 9) argues that it is impossible to separate these functions as ‘truth and value are things of this world...they do not escape social constraints’. Miskovic-Lukovic’s (2009) paper on hedges such as *‘kinda’* and *‘sort of’* argues that they perform a range of different function from being fillers that ‘allow us to think of what next to say, or just to indicate that we intend to go on talking’ (Leech and Svartvik, 1996:11) to ‘adverbials whose function is to reduce the force’ Miskovic-Lukovic (2009: 603). As discussed throughout this section it is difficult to determine whether or not hedges are functioning from an interpersonal perspective. However there is greater consensus with regards to the way in which they act to weaken an evaluation and this is broadly the way in which they are analysed in this thesis. This is discussed in greater detail in section 3.2.3

2.2.4 Evidentiality

Evidentiality in its broadest sense can be understood as anything that involves attitudes towards knowledge (Chafe 1986). In this sense it shares a similarity with the notion of stance as discussed above, in particular, the sub-class of epistemic stance. Dendale and Tasmowski (2001: 340) suggest that reference to the source of information, evidentiality in the narrow sense, has been closely linked to markers of certainty about information because ‘... the linguistic markers encoding these two semantic domains are often the same’. They then go on to state that the exact nature of the relationship between these two domains is a significant problem in this research area. While it should be clear that indicating the source of information is not necessarily to comment on its reliability, this distinction is not always clear from the terminology. The relationship between evidentiality and modality is also imprecise in nature, with three different relationships proposed in the literature:

- Disjunction – where they are conceptually distinguished from each other
- Inclusion – where one falls within the semantic scope of the other
- Overlap – where they partly intersect

Hardman (1986: 115) defines evidentials as indicating ‘how one has knowledge of what one is saying’ and this is an example of what Willet (1988: 54) terms ‘evidentiality in the narrow sense’. That is, it denies any explicit relationship between evidentiality and modality. The more prevalent view is that of inclusion where one of the concepts is regarded as falling within the scope of the other. In this case evidentiality is then used to refer both to the source and the reliability of the speaker’s knowledge. This view is proposed by Matlock (1989:215) who states ‘Evidentials, linguistic units comprising part of epistemic modality, code a speaker’s source of information, and some degree of certainty about that information.’ Willet (1988) also suggests that evidentiality is primarily modal. Evidentiality is more often included in modality and this is perhaps because evidentiality can be used to help define epistemic modality (Dendale and Tasmowski 2001). The final view is that of Van der Auwera and Plungian (1998: 86) who suggest that ‘an overlapping relation can be found where modality and evidentiality intersect.’ This overlap is then occupied by what they term ‘inferential evidentiality’ which they suggest is the same as the modal value of epistemic necessity.

Having looked at the different ways in which evaluations can be grammatically framed to strengthen or weaken a proposition, it is important to note that any perceived strengthening of a proposition is inherently weaker than simply stating something as fact. Halliday (1994) argues that even a high value modal ‘certainly’ or ‘always’ is less determinate than a polar form: *that’s certainly John* is less certain than *that’s John*. In other words, a speaker only has the need to say that they are certain when they are not. This is a view that White (2003) shares, arguing that in contrast to representing something as fact, high certainty can be seen as a form of hedging.

As has been discussed throughout section 2.2, there are significant areas of overlap between evaluative realisations analysed under the differing terms of stance, modality, hedging and evidentiality. Although the exploration of evaluation in text has often been focused on specific markers, it is widely accepted that ‘while evaluation may be concentrated at particular points or phases in the text, it is nonetheless encoded throughout texts through the deployment of a wide range of linguistic resources’ (Hood 2004: 44). Indeed Volosinov (1973: 105) goes further stating that:

‘No utterance can be put together without value judgement. Every utterance is above all an evaluative orientation. Therefore, each element in a living utterance not only has a meaning but also has a value’.

In addition, while certain parts of a text may appear to be free from evaluation, Macken-Horarik (2003) suggests that ‘spans of text not marked explicitly for attitude can carry evaluative meaning by virtue of their cohesive links to other more attitudinal parts of the text’. This again highlights the need for an analytical approach that attempts to account for all the evaluative content of a text, rather than just looking at surface level features. So despite arguing that taking a separating approach can lead to problems of clarity, Martin’s (2003) work on Appraisal theory strikes a middle ground with Thompson and Hunston (2003: 4) stating that he ‘follow this separating approach, but expands the account of attitudinal meaning greatly’. This body of work (Martin 1992, 1995, and 2003), White (1998, 2003) and Martin and White (2005) is something that I return to in Chapter 3. Section 2.3 will now look at how meanings in individual lexical items have been explored to aid understanding in evaluation.

2.3 Lexis

Hunston and Thompson (2003: 14) argue that ‘Some lexical items are very clearly evaluative, in the sense that evaluation is their chief function and meaning.’ Evaluative lexical items can be found across many parts of speech:

- Adjectives: Splendid, terrible, surprising, obvious, important, possible, untrue
- Adverbs: Happily, unfortunately, plainly, interestingly, possibly, necessarily
- Nouns: Success, failure, tragedy, triumph, likelihood
- Verbs: Succeed, fail, win, lose, doubt

For words that are so obviously and readily evaluative in their nature, there is perhaps less need for analysis in close detail as the attitude and evaluative stance of the speaker can be 'read' more easily. However, not all lexical items are so easily identified as 'evaluative' or 'non-evaluative'. In many instances a word may have no obvious evaluative content and its polarity may therefore be determined by the context in which it is used. Hunston and Thompson (2003) use the word '*student*' to exemplify this point, suggesting it could take one of three evaluative positions:

- A neutral classification of those who are studying
- A positive evaluation associated with a desire to learn and to better oneself
- A negative evaluation associated with laziness and untidiness

They propose a corpus based technique that allows a researcher to study large amounts of naturally occurring language. This approach has led to developments in the field of what has been termed Semantic Prosody, which is discussed below in Section 2.3.1.

Channell's research into connotation is also concerned with '...where it [attitude] is carried by individual lexical items, or by semi-fixed expressions, rather than on examples where the function is carried by whole sentences or stretches of text' (Channell 2000: 39). Her research is based on words that encode evaluation as part of their meaning, alongside other features, rather than those whose overt and only purpose is to evaluate. One of the main findings to come out of this research is that 'important aspects of the use of lexical items are not open to conscious reflection, particularly when these concern something as important to meaning as positive versus negative orientation' (Channell 2000: 54). So while it is apparent that evaluation, to some extent at least, can be encoded and realised through individual lexical items, it is also necessary to look at how the relationship between lexical items can create new meanings.

2.3.1 Semantic Prosody

Semantic prosody as a concept was first suggested by Sinclair (1991) and subsequently taken up by Louw (1993). As a result of developments in corpus linguistics it became possible to study vast numbers and patterns in language usage. Corpus linguistics allows researchers to observe language as it is used and to analyse it as raw data. Each data item can then be analysed making it possible to spot collocation patterns in the surrounding language. By looking at each example of the word under study, and the surrounding language, it is possible to see whether it collocates predominantly with positive or negative utterances. Louw (1993: 50) defines Semantic Prosody as ‘...a form of meaning which is established through the proximity of a consistent series of collocates, often characterisable as positive or negative’. Examples that he provides include the semantic prosody that is attached to apparently neutral words such ‘utterly’ and ‘symptomatic’. Using this corpus methodology Louw concordanced ‘utterly’ and ‘symptomatic’ and found that their collocates tended to be negative. He argues that ‘the concordance shows that *utterly* has an over-whelmingly ‘bad’ prosody’ and that in the case of *symptomatic* there is ‘overwhelming evidence of a negative prosody in this concordance’(Louw 1993: 171).

Channell (2000: 38) explains it thus:

‘A given word or phrase may occur most frequently in the context of other words or phrases which are predominantly positive or negative in their evaluative orientation. As a result, the given word takes on an association with the positive, or more usually, the negative, and this association can be exploited by speakers to express evaluative meaning covertly.’

This suggests that lexical items which are not obviously evaluative may still have a positive or negative polarity, allowing the speaker to realise their evaluations covertly.

Stubbs (2001: 6) suggests that 'a major finding of corpus semantics is that words and phrases convey evaluations more frequently than is recorded in many dictionaries.' The potential to make negative evaluations in such a covert way is also recognised by Louw (1993: 173) who suggests that semantic prosodies have 'in large measure and for thousands of years, remained hidden from our perception and inaccessible to our intuition.' Although their existence has only been highlighted relatively recently, Breal (1897) seemed to be alluding to this process nearly a century earlier when he referred to 'transference of meaning' as a result of habitual collocation, a phenomenon he termed 'contagion'. The importance of looking at lexical items in context was also proposed early on by Firth (1935: 37). He argued that 'the complete meaning of a word is always contextual and no study of meaning apart from a complete context can be taken seriously'.

More recently the problems associated with analysing a word's meaning in isolation have been discussed by Sinclair (1994: 23). He states that 'the meaning of words chosen together is different from their independent meanings' and uses the term 'delexicalisation' to explain the loss of information and meaning that occurs when words are looked at in isolation. This means that an analysis based only on a word in isolation will not provide a full account of a word meaning.

Louw (1993: 171) argues that semantic prosody appears to be more common in negative attitudes and evaluations. 'There seem, *prima facie*, to be more 'bad' prosodies than 'good' ones, but the latter certainly exist and the principles on which they work are the same'. One reason for this is that it could be used as a face saving act as it is harder to make negative comments. Stubbs (1995) established that the verb '*cause*' collocates overwhelmingly often with words for unpleasant events such as war, famine or disease but rarely, if at all, with positive events or actions. He argues that 'a true definition of the word should not be 'make something happen' but 'make something bad happen' (Stubbs 1995: 28). Semantic prosody, therefore, can be seen to offer an indirect and implicit way of making a negative evaluation.

However, whilst it has been acknowledged that semantic prosody 'has become one of the more important concepts in corpus linguistics' (Whitsitt 2005: 283), it has also been questioned and critiqued. Coffin and O' Halloran (2006) use a corpus based technique to explore the phenomenon of 'dog whistle- politics' and covert evaluation in newspaper articles. 'Dog-whistle politics' is the idea that a target audience may come to interpret seemingly neutral or positive representations of groups or people in negative ways. In the same way that a dog-whistle can be heard by dogs but is pitched too high for the human ear, a politician or journalist may pitch a message 'high enough' for their intended reader to hear it, without it appearing explicitly negative to an unprepared hearer.

The ability of a reader to tune in to this 'pitch' is dependent on the degree to which a target reader has been primed and positioned through prior reading. To explore this notion, a corpus of previous articles was collected by Coffin and O' Halloran (2006) in addition to the main article under analysis. This is part of what Halliday (1992, 1993) and Halliday and Matthiessen (1999) refer to with the term '*logogenesis*', which refers to the dynamic build-up of meaning as a text unfolds. Coffin and O' Halloran (2006: 81) suggest that their method is doubly logogenetic as it takes into account not only the build-up of meaning within an individual text, but also within a preceding corpus of texts.

While this creation of meaning through repetition and collocation seems to be comparable to the notion of semantic prosody, Coffin and O' Halloran suggest that rather than being a result of the construction of '...a consistent aura of meaning with which a form is imbued by its [immediate] collocates' (Louw 1993: 157), it is instead more a result of words and phrases being 'pre-contextualised'. They argue that as concordance lines only reveal five lines of co-text, the remainder of the text, and other related texts, are left out of the meaning-forming process for the analyst. They also suggest that the concept of semantic prosody is not fine-grained enough and that they 'did not find clear instances of semantic prosodies backed up by sufficient evidence' (Coffin and O' Halloran 2006: 91).

Whitsitt (2005) is even more vociferous in his criticism. He argues that there is little, if any distinction between what has been termed Semantic Prosody and the pre-existing notion of connotation. 'Semantic Prosody is simply connotation spread over several words, and connotation is semantic prosody that no longer shows how the process of semantic transfer takes place' (Whitsitt 2005: 285). He also questions why some words would take on meaning from their collocates while others would not. Sinclair (1996: 80) argues that there can be 'a frequent co-occurrence of words which does not have a profound effect on the individual meanings of the words' but Whitsitt (2005: 289) suggests that the main problem is that the notion of Semantic Prosody assumes that there '...are some words which are full and others empty and that when a full word is next to an empty word, it seems unable to not 'pour' its meaning into the empty innocent one, which in turn seems unable to refuse'. This dispute over where meaning resides, *in* words or *between* words, can be seen as the difference between the 'open choice principle' and the 'idiom principle' (Sinclair 1991: 109). Whitsitt (2005: 289) describes this as being the difference between individual words being seen as '...independent, individual units, and not bound in chunks of language', and '...the idea that words bond with each other, and tend to go together'.

While there is some dispute as to the specifics regarding semantic prosody and connotation, it is apparent that meaning, and therefore evaluation, can be covertly realised and encoded in words that might at first glance appear neutral. This section argues that in addition to individual lexical items whose evaluative meaning is obvious and uncontentious, there are also words that have accrued some of their meaning, particularly their polarity, from their collocation with other words. Once this new meaning has been established, the individual word will now be permanently 'marked', even when in not in the company of the words that have led to the new meaning. Having looked at the way in which meaning and evaluation can be determined and embedded within individual lexical items, section 2.4 will look at the way in which research from the field of Phonology and Non Verbal Communication could also be used to aid understanding of evaluative talk.

2.4 Phonology and Non Verbal Communication

Phonology is the study of the way in which sounds are organised and the way that they function. In this instance it is the meaning, specifically the emotion or attitude that is encoded in the variations of these sounds, that is of interest. The most common areas of study are those of pitch, tone and pace and these are discussed below. Non verbal communication encompasses body language; facial expressions, and arm and hand movements and each of these will also be introduced and discussed in relation to analysing emotions and attitudes on evaluative talk.

2.4.1 Prosody: Pitch, Tone and Pace

The field of prosody is concerned with variations in a speaker's pitch, tone and pace. The importance of understanding prosody is highlighted by Buchanan et al (2000: 1) who propose that 'the term prosody describes the non-propositional cues that may be passed along through language, including intonation, stresses, and accents. Emotional prosody involves the expression of emotion through the intonation of spoken language'. The view that the analysis of prosody can be used to interpret emotion and affect is supported by Mozziconacci (2001: 1) who states that '...prosody not only carries information on word stress, phrasing and emphasis, but is additionally thought to be strongly related to speaker specific characteristics, and factors such as the expression of the speaker's emotions and attitudes'. There is much research proposing a link between the expression of emotion and attitude and variability in parameters such as voice quality, volume, rhythm, pitch, (for example Carlson, 1991; Leinonen et al, 1997 and Protopapas & Lieberman, 1997). Mozziconacci's (2001: 31) research has shown that '...intended emotions were recognized far above chance level by the subjects'. This follows on from previous research by Siegwart and Scherer (1995) which reported that in studies where the participants' task was to infer the underlying emotion by listening to natural speech, the accuracy was approximately five times higher than the level of chance. Though it is clear that a prosodic analysis could add detail to this research, Mozziconacci (2001: 1)

argues that 'quantitative details of the correspondence between prosodic features and emotion or attitude, are still poorly understood'.

2.4.2 Non Verbal Communication

Much early work on communication was focused not on the language that was used but rather on the *way* in which it was said. In addition to this, the way in which gestures and facial expressions were used was also given priority. The most influential early work, into the study of body language, was perhaps that of Mehrabian and colleagues (1967; 1969, 1971) and Birdwhistell (1952; 1970, 1971). Their studies suggested that the total impact of a message is 7% verbal (words only) 38% vocal (including tone of voice, inflection and other sounds) and 55% non verbal.

On the face of it, this breakdown of interpersonal communication should be particularly relevant to this research as Mehrabian et al (1967: 35) state that these findings were based on 'communications of feelings and attitudes (i.e., like-dislike)'. However there have been many criticisms of this research particularly with regard to the experimental design and the way in which results were then determined. The main point being that the results were arrived at through the amalgamation of two separate studies with different research questions, methodologies and participants. There has also been criticism towards the over-simplification and misinterpretation of the main findings and the suggestion that they can be applied to all human communication. Indeed, Mehrabian himself has acknowledged that 'these findings regarding the relative contribution of the tonal component of a verbal message can be safely extended only to communication situations which match the experimental design' (Mehrabian & Wiener 1967: 113-114).

It has also been argued that the positioning and movement of a speaker's arms and hands can be used to realise attitudinal positions (Müller 2004 and McNeill 2005). This use of the hands and arms in communication is ubiquitous, and McNeill (2000: 1) argues that 'in almost any language and under nearly all circumstances, you will see what appears to

be a compulsion to move the hands and arms in conjunction with the speech'. However Krauss et al (1991) suggest that there is little evidence to support the notion that hand gestures are primarily used to aid communication. They go on to suggest that all studies on gesture have '...found a considerable amount of gesturing when speaker and listener could not see each other, something that is difficult to square with the 'gesture as communication' hypothesis' and conclude that 'it may also be the case that the contribution gestural information makes to communication typically is of little consequence' (Krauss et al 1991: 745).

So although it is apparent that gestures and hand movements do play some part in the communication of attitudes and emotions and indeed in communicative effectiveness in general, the level to which they do this is still somewhat unclear. It is also accepted that body language and facial expression, as well as prosody, do have a role in the expression and realisation of emotion and affect. However, the notion that linguistic content is responsible for as little as 7% of meaning is strongly contested. The general mood of a speaker may be accurately assessed by analysing both their nonverbal communicative actions and the prosodic nature of their speech. However, the subject matter that is both the cause and the target of the attitude or emotion can only be understood by looking at the language and words that are used. Therefore a close analysis of word choice (lexis) and construction (grammar) is arguably the best way of not only assessing the prevailing attitude or emotions but also the detail of towards what or whom they are directed. In addition to the doubts as to the merits of analyses based on prosody and body language, the nature of the data collected in this thesis means that it will not include analysis of prosodic or non verbal features.

This chapter set out to explore evaluation from the different perspectives of Discourse, Grammar, Lexis and Phonology and Non verbal communication. Having discussed the literature across these different fields, Chapter 3 will now introduce two theoretical frameworks that have been developed to aid the analysis of discourse in general. Their suitability for specifically analysing evaluative talk is also assessed.

3. Theoretical Frameworks for the Analysis of Evaluation

In this Chapter I will discuss two frameworks that can be used to analyse evaluation in discourse. These frameworks are, Systemic Functional Linguistics and Appraisal Theory. Each of these will be introduced and then assessed in terms of their usefulness for helping to measure a product's affective impact.

3.1 Systemic Functional Grammar

Systemic Functional Grammar is synonymous with the work of Halliday (1994: F40) and is based on the premise that 'each element in a language is explained by reference to its function in the total linguistic system... a functional grammar is one that construes all the units of the language – its clauses, phrases and so on – as organic configurations of functions'. This approach to grammar and language can be used to understand and evaluate the impact of different language choices in the construction of different communications. In common with other descriptive grammars, it is an attempt to understand and map the way language has evolved and is actually used, rather than stating what correct or incorrect usage is. In this aspect it is very different from more prescriptive or corrective grammars. The systemic nature of this framework means that whatever is chosen in one system forms the way into a set of choices in another, and this process can be continued for as long as is needed to understand the text under analysis (Halliday 1994).

Within Systemic Functional Grammar, meaning can be seen as composed of three metafunctions; the Experiential, the Textual and the Interpersonal. For this research, I am primarily interested in the Interpersonal and Experiential metafunctions as it is the relationships between the speaker and the object that is being explored. The Interpersonal metafunction is related to the ways in which the speakers in a communication interact, and specifically looks at the '... aspects of the grammar which can be identified as enabling us to interact through language' (Thompson 1996: 38). These aspects include the different speech roles that are taken by the communicators and

allow the language to be viewed in terms of its purpose as a communicative event. Each clause functions as either a demand or an offer for either information or for what Halliday terms 'goods and services'. This function constitutes the mood of the clause and reveals the speaker's objectives, and also how they expect the addressee to respond. The choices available with regards to different speech roles are demonstrated in Figure 7.

Role in exchange ↓	Commodity exchanged →	(a) goods-and-services	(b) information
(i) giving		Offer I'll show you the way	Statement We're nearly there
(ii) demanding		Command Give me your hand	Question Is this the place?

Figure 7 Basic Speech Roles

The experiential metafunction is a set of resources for looking at how different entities relate to the world and each other. This allows language to be analysed in terms of who is doing what to whom and also in what way they are going about it. There are different ways of representing the same actions and the choices that are made are often informed by the ideology of the speaker. An analysis of these two metafunctions could be used to provide some insight into consumers' opinions and attitudes about particular products, both in relation to themselves and to the world in general. However, although it would be possible to use this framework, there are two main reasons why it might be more productive to look elsewhere for the main theoretical framework that will be applied to the data collected in this research. These are explained below in section 3.2.

The second frameworks that I will assess is that of Appraisal Theory. This theory can be viewed as an extension and broadening out of the Interpersonal sphere of Systemic Functional Grammar (Halliday 1994).

3.2 Appraisal Theory

Appraisal Theory is a system of analysis that has been developed by a large research group over the last fifteen to twenty years. The principal contributors and proponents of this framework are Martin (1992, 1995, and 2003) and White (1998, 2003). As with all theories, it is not closed or complete and amendments and updates are continuously being proposed and considered. Coffin and O'Halloran (2006: 84) state that 'throughout the last decade, appraisal categories have been put to the test in numerous contexts and, as a consequence, modifications (an ongoing process) have been made'. Many of these are discussed through the online Appraisal Discussion Group to which Martin and White both contribute regularly.

Appraisal Theory was borne out of necessity as a result of a research project that ran throughout the eighties and nineties. Martin (2003: 171) states that 'at about this time, a group of functional linguists in Sydney began work on developing a comprehensive framework for analyzing evaluation in discourse.' The project which necessitated this change was the Write it Right project which was set up to '...explore the literacy requirements of the discourses of science, technology, the media, history, English literature studies, geography and the visual arts' (www.grammatics.com 2005). Although work in this field had started to progress and develop, Martin (2003: 171) suggests that 'A good deal of the political impetus and funding for this work arose in the field of educational linguistics and the development of Australia's genre-based literacy programs'.

The starting point of the research was set within the frameworks of the Interpersonal systems of Systemic Functional Grammar. However, it soon became apparent that there was a need to revise or at least broaden some of the existing descriptive frameworks to incorporate areas which had not been sufficiently addressed in the existing literature. As a result, this research '...gave rise to an account of JUDGEMENT as a set of meanings by which speakers appraise the behaviour of human individuals and to APPRECIATION as a

set of meanings for making aesthetic and related assessments of objects and products' (Martin 2003: 35). These terms are introduced and discussed in detail below.

Appraisal Theory makes use of three sub-systems that can be applied to texts to analyse them in terms of different areas of interest. The three areas are ATTITUDE, ENGAGEMENT and GRADUATION. The following sections will briefly describe each of these three areas and assess their potential for helping to analyse the data that has been collected for this study. Although I argue that the framework maps onto areas that are likely to be of interest to the market researcher, there are also some areas where small amendments or adaptations are required. These are also discussed below.

3.2.1 Attitude

Of the three different areas that together constitute Appraisal Theory, the most useful tools for the analysis of market research interview transcripts may lie within the ATTITUDE framework as this is specifically intended to help assess:

- i. How a speaker is emotionally disposed to the subject of the communication
- ii. How the subject of the communication compares to accepted norms and values
- iii. How the subject of the communication creates an impact on the speaker in terms of form, appearance and aesthetics.

Painter (2003: 184) describes ATTITUDE as 'a domain concerned with the linguistic expression of positive and negative attitudes'. This fits in very well with the demands and requirements that need to be met to allow the data to be analysed most effectively.

These three further sub-categories are labelled AFFECT, JUDGEMENT and APPRECIATION and are looked at in further detail below.

3.2.1.1 Affect

Martin (2003: 145) describes AFFECT as a resource for ‘construing emotional responses’ and argues that it can be realised across a range of different terms which are shown in Halliday’s (1994) terms in Table 2.

Quality	Describing participants	a happy boy	Epithet
	Attributed participants	the boy was happy	Attribute
	Attributed to participants’ manner	the boy played happily	Circumstance
Process	Affective mental	the present pleased the boy	
	Affective behavioural	The boy smiled	
Comment	Desiredative	happily, he had a long nap	Modal Adjunct

Table 3 The different ways in which Affect can be realised (Martin 2003)

In addition to the different ways in which AFFECT can be realised, it can also be further analysed as either positive or negative AFFECT. This is assessed on the basis of whether or not the feelings are culturally understood as good or bad. Therefore ‘*the boy was happy*’ would be positive AFFECT and ‘*the boy was sad*’ would be assessed as negative AFFECT. Although there are some situations where it might be considered good to be sad, to release emotions rather than bottling them up perhaps, it is the standard cultural reading of the emotion that is taken as the assessed value. As well as the direction of the emotion, its strength should also be considered. In terms of intensity, AFFECT can be graded as being towards the higher or lower end of the scale. However, while the grading cannot be considered to be highly defined, there is an inherent difference in strength between ‘*I like that*’, ‘*I love that*’ and ‘*I adore that*’. The different ways in which AFFECT can be invoked can be understood in terms of being a surge of emotion, a predisposition to a certain way of thinking or as a continuous mental state. These differences map onto behavioural ‘*she frowned at him*’, mental ‘*she disliked him*’ and relational ‘*she felt cross with him*’ processes (Martin 2003). The AFFECT branch of the framework is split into the

four sub-branches of HAPPINESS, SECURITY, SATISFACTION or DESIRABILITY with each of these having a positive or negative split. Martin and White (2005: 49) describe AFFECT as grouping ‘emotions into three major sets having to do with UN/HAPPINESS, IN/SECURITY and DIS/SATISFACTION’. The category of un/desirability was added at later date (Krsner 2000).

The UN/HAPPINESS set relates to feeling happy or sad in relation to a trigger. This covers the range of feelings that ‘are probably the first to come to mind when we think of emotions’ (Martin and White 2005: 49). IN/SECURITY relates to feelings of peace and anxiety while DIS/SATISFACTION ‘deals with our feelings of achievement and frustration in relation to the activities in which we are engaged, including our roles as both participants and spectators’ (Martin and White 2005: 50).

Affect Type	Surge of Behaviour	Disposition
Happiness	Laugh, hug, rejoice	Cheerful, love, like
Unhappiness	Cry, abuse, revile	Sad, dislike, hate
Security	Assert, proclaim, entrust	Confident, assured, trusting
Insecurity	Restless, twitching, faint	Uneasy, anxious, staggered
Satisfaction	Compliment, industrious, attentive	Absorbed, impressed, satisfied
Dissatisfaction	Fidget, yawn, caution	Stale, bored with, angry

Table 4 Examples of Affect across the three main categories of emotion (Martin 2003)

This table shows examples of AFFECT across each of three main emotional categories in addition to showing the difference between how a surge of behaviour or general disposition might be realised in language. Painter (2003) suggests that AFFECT is also institutionalised within both JUDGEMENT and APPRECIATION, as explained below:

An attitudinal word or other utterance need not directly express the feelings of a speaker; instead it may construe someone’s behavior in positive or negative terms within a framework of social and ethical values (the system of JUDGEMENT), or it may evaluate their artistic,

intellectual, sporting, professional or other products and processes (the system of APPRECIATION). Painter (2003: 189)

Although each of these categories is introduced as a separate part of the framework, there are frequent sites of overlap and ‘fuzzy boundaries’.

3.2.1.2 Judgement

JUDGEMENT is related to human behaviour and how it compares to social norms. Krsner (2000: 93) states that ‘JUDGEMENT’ constitutes the semantic resource for construing evaluation of behaviour in the context of institutional norms about how people should or should not behave’. JUDGEMENT is split into five further categories:

- NORMALITY
- CAPACITY
- TENACITY
- VERACITY
- PROPRIETY

NORMALITY relates to how special a person is and how their behaviour relates to these social norms. Adjectives such as *unfortunate, peculiar, average* can be accounted for here. There seems to be some overlap between NORMALITY and VALUATION. However, in VALUATION, ‘*average*’ or ‘*strange*’ refers to the overall evaluation of whether a product works, whereas for NORMALITY it refers to the person and whether or not they are special or out of the ordinary. CAPACITY covers formulations relating to how capable the person is. *Balanced, sensible* and *together* would all be tagged in this way. Again there is the potential for overlap with COMPOSITION: BALANCE, as introduced below with the distinction being made dependant on whether the recipient of the evaluation is a person and their behaviour or an object. Both VALUATION and COMPOSITION relate to APPRECIATION which is discussed below in section 3.1.1.3.

TENACITY accounts for the speaker's commitment and dependability. Krsner (2000: 94) writes that being 'brave, heroic, energetic and so forth is associated with a positive disposition while lazy, unreliable, apathetic and so forth encode a negative evaluation'.

VERACITY relates to the truthfulness of the person and how honest they are; *truthful, candid* and *deceptive*. PROPRIETY deals with how ethical the behaviour is; *kind, respectful* or *corrupt*. These five categories are grouped together under the headings of SOCIAL ESTEEM and SOCIAL SANCTION. NORMALITY, CAPACITY and TENACITY combine to account for evaluations related to social behaviour and VERACITY and PROPRIETY combine to account for the 'domain of moral regulation' Krsner (2000: 95).

JUDGEMENTS TO DO WITH SOCIAL ADMIRATION OR CONTEMPT		
SOCIAL ESTEEM	POSITIVE (ADMIRE/BE CAPTIVATED BY)	NEGATIVE (HOLD IN CONTEMPT/PITY)
Normality (Fate) (usuality)	normal, outstanding, lucky, remarkable	peculiar, odd, eccentric, unlucky, abnormal
Capacity (ability)	competent, powerful, witty	weak, incompetent, stupid, foolish, incapable
Tenacity (Resolve) (inclination)	plucky, heroic, curious, resolute, self-reliant	cowardly, rash, apathetic, obstinate, vexatious, lazy, servile, complacent

JUDGEMENTS TO DO WITH MORAL RIGHT AND WRONG		
SOCIAL SANCTION	INSTITUTIONAL PRAISE	INSTITUTIONAL BLAME
Veracity (Truth) (probability)	honest, frank, real, genuine, credible	deceitful, fake, bogus, dishonest, deceptive
Propriety (Ethics) (obligation)	right, good, ethical, kind, generous, loyal, forgiving	wring, evil, sinful, mean, cruel, greedy, arrogant, corrupt

Table 5 System of Judgement (Martin and White 2003: 53)

In line with both AFFECT and APPRECIATION, JUDGEMENT has both positive and negative dimensions and like AFFECT it is assessed on the basis of whether or not the evaluations

are culturally understood as good or bad traits or behaviours. In terms of applications to my research, this is very much field dependant as the different requirements placed on different products will mean that in some instances it may be desirable for an item to be 'strong' whereas in others being 'weak' may be the preferred state. Therefore to help judge the direction of the evaluation, further techniques will be used. These are outlined below. As shown in the Table 5 above, JUDGEMENT is primarily concerned with human behaviours and actions. Therefore occurrences, as defined by this framework, are likely to be infrequent within the transcripts collected for this pilot study, as the subjects of the communication are objects and concepts rather than humans and their behaviour. However, it is possible to use this framework to assess respondents' intuitions about how representative an example or stimulus is of their expectations or previous experiences of a product category. That is, given that they are likely to have frequent interactions with deodorant cans and packaging, they are likely to have a sense of what is or is not 'normal'. This will be tagged under the JUDGEMENT: NORMALITY branch of the framework, as shown in the top line of Table 5 Although this is designed to account for human behaviour, I argue that it can be adapted to account for formulations where the speaker discusses how much an object matches up to their expectations of what it should look like.

3.2.1.3 Appreciation

APPRECIATION is related to the impact that a product or object has on an individual so this appears to be the quality most closely related to the areas which might be of interest to market researchers. Martin and White (2005: 56) suggest that APPRECIATION accounts for 'meanings construing our evaluations of "things" especially things we make'. The framework for analysing APPRECIATION can be used to assess what the speaker thinks about the product in terms of their overall reaction to it, their thoughts on its composition and whether or not they like the final design and think it was a worthwhile enterprise. Page (2003: 214) describes it as being to do with the 'aesthetic attributes associated with an entity'. This is obviously extremely useful for research that is looking to assess the ways in which the designs of products can positively affect the

chances of a consumer buying them. The framework of APPRECIATION is organised around the three variables:

- REACTION
- COMPOSITION
- VALUATION

Each of these has subsequent sub-branches on which evaluative utterances can be placed.

REACTION deals with both the IMPACT of the product and its perceived QUALITY. Martin (2003: 160) describes IMPACT as having to do with 'the degree to which the text/process in question captures our attention'. Being able to assess the immediate impact of a product is very important and this ties in with looking at ways of making the 'first moment of truth' more appealing to consumers. Krsner (2000: 97) defines QUALITY as accounting for 'assessments of an object's likeability'.

COMPOSITION is related to the BALANCE and the COMPLEXITY of an object. COMPOSITION: BALANCE relates to 'perceptions of proportionality' (Martin 2003: 160) and accounts for how the speaker feels about the physical aspects of the target of the communication. COMPOSITION: COMPLEXITY is concerned with the details of the object and how easy it is to understand or use. An analysis of COMPOSITION might help to understand how a respondent feels about the stimulus in terms of how well it fits together and all aspects of its size, in addition to assessing how easy it is to use or to understand. This might be of particular interest to the designers or brand owners of a product which has an unfamiliar feature such as a new opening method which needs to be easily understood.

VALUATION simply looks at whether the overall design works and is worthwhile. Krsner (2000: 98) describes it as relating to 'the social significance of a product or phenomenon'. This can be related to the 'second moment of truth' as it is almost an after the event evaluation of the product once it has been looked at and experienced by the consumer or respondent. In addition to this, all three areas of APPRECIATION have both a

positive and negative dimension mirroring the possible positive and negative evaluations of texts, objects and people. These three variables, in both their positive and negative manifestations, provide a strong framework for assessing and understanding the different ways in which people make judgements and evaluations.

	Positive	Negative
REACTION: IMPACT <i>(Did it grab me?)</i>	Arresting, captivating, engaging, fascinating	Dull, boring, tedious, uninviting
REACTION :QUALITY <i>(Did I Like it?)</i>	Lovely, beautiful, appealing, splendid	Plain, ugly, repulsive, revolting
COMPOSITION: BALANCE <i>(Did it fit together?)</i>	Balanced, harmonious, proportional	Unbalanced, discordant, distorted, stretched
COMPOSITION: COMPLEXITY <i>(Was it hard to understand)</i>	Simple, elegant, intricate, detailed	Confusing, monolithic, extravagant
VALUATION <i>(Was it worthwhile?)</i>	Original, unique, challenging	Insignificant, shallow, reactionary

Table 6 A Framework for analysing APPRECIATION (Martin 2003: 160)

Table 6 shows the trigger questions that can be used to assess which branch of the APPRECIATION framework an evaluation should be placed under. It also provides single lexical items that are representative of each category. As well as single items, multi-word phrases or expressions can also be tagged for each of these categories. As Page (2003: 221) states, 'APPRAISAL might be realized by a single word, phrase or whole proposition'. The question of how APPRAISAL is realised across differently sized units is also addressed by Hunston (2003) and Macken-Horarik (2003). Macken-Horarik (2003: 317) also proposes four 'environmental frames' that should be considered when assessing the size of an evaluative item.

- a) The co-text surrounding any evaluative item.
- b) The global frame of the text itself.

- c) The intertextual frame of other texts conditioning the production and reading of the text.
- d) The contratextual frame of 'reading against the grain'.

In addition to exploring the size of an evaluative item, these different spheres or frames are also likely to help determine which evaluative tag is applied. Macken-Horarik (2003: 317) notes that "coding" of lexical instantiations of appraisal needs to be conditioned by a consideration of their place within these larger co-textual environments'. Although Martin (2003: 161) suggests that 'what counts as appraisal depends on the field of discourse', Painter (2003: 202) argues that APPRECIATION is the most context dependant: 'APPRECIATION is the domain of attitude whose parameters are most contingent upon the particular fields under attention'. A more detailed discussion of the way in which the context can influence the tagging is carried out in section 5.4.1.1.4, Tagging ENGAGEMENT.

The three sub-branches of AFFECT, JUDGEMENT and APPRECIATION are grouped together under the heading of ATTITUDE and White (1998) argues that to some extent all three represent some kind of emotional response. JUDGEMENT is the formulation of feeling with regards to human behaviour and APPRECIATION is the formulation of feelings with regards to products and processes. As such the boundaries between the categories detailed above are fuzzy and at times require subjective decisions with regards to analysis. The interconnectedness between APPRECIATION, JUDGEMENT and AFFECT is shown below in Figure 8. Though JUDGEMENT and APPRECIATION are described as discrete categories, there is some level of overlap through the AFFECT branch of the framework, and this can be seen to straddle these two other categories.

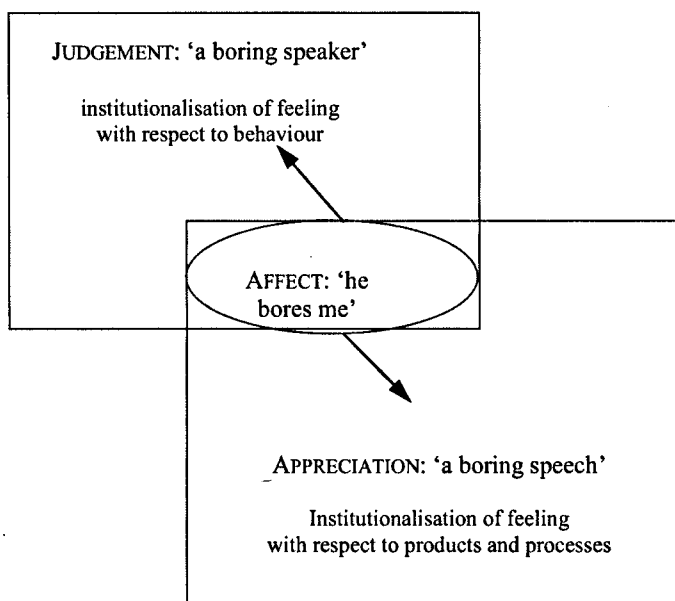


Figure 8 The affectual basis of judgement and appreciation (White 1998: 154)

Having looked at where in the framework an evaluation might be placed, the following two sections will look at the different ways in which these evaluations can be strengthened or weakened and how these approaches can be categorised.

3.2.2 Engagement

'ENGAGEMENT' is concerned with looking at how ideological positions are expressed inter-subjectively and encompasses the exploration of words or phrases traditionally described under the categories of modality, hedging and evidentiality. The literature of these areas is split into two distinct camps; those that follow the dialogic approach proposed by Bakhtin (1981, 1986) and Volosinov (1973, 1995) and those that assert that these words and phrases have the sole function of indicating the speaker's commitment to the truth value of the proposition. Research into modal auxiliary verbs has previously focused on issues related to the levels of certainty, obligation, or probability that are assigned to an event or opinion (Lyons 1977, Palmer 1986 and Chafe and Nichols 1986). However, as noted in section 2.2.2, Martin and White (2005) and White (2000, 2003) follow Bakhtin and Volosinov's influential view of verbal communication. Bakhtin (1981, 1986) and Volosinov (1973, 1995) contend that all

verbal communication is 'dialogic' in that any speech necessarily both refers to what has been said before and anticipates the response of actual, potential or imagined readers or listeners. As Volosinov states:

'The actual reality of language-speech is not the abstract system of linguistic forms, not the isolated monologue utterance, and not the psychological act of its implementation, but the social event of verbal interaction implemented in an utterance or utterances...it responds to something, affirms something, anticipates possible responses and objections, seeks support and so on.'

(Volosinov 1995:139)

Bakhtin argues that this relationship between seemingly unconnected discourses is a continuous cycle as 'Any utterance is a link in a very complexly organized chain of other utterances' (Bakhtin 1986: 69). Martin and White (2005) agree with this viewpoint and suggest that rather than showing the level of engagement that the speaker has with their own evaluation, Appraisal Theory instead indicates the speaker's level of engagement with all possible readers or listeners. White (2003) sets out a taxonomy of the different ways in which intersubjective positions can be adopted or framed and proposes that these resources can be classified as either 'dialogically expansive' or 'dialogically contractive'. This simplistic distinction is something that will be returned to in section 5.4.1.1.4 with regards to the discussion of the way in which ENGAGEMENT has been tagged in the main study. The difference between 'dialogically expansive' and 'dialogically contractive evaluations' is perhaps best understood in terms of the difference between stating a fact and offering an opinion. Where a speaker frames their utterance in a way that makes '... a claim about some observable, verifiable state of affairs in the *experiential* world' (White 2003: 264) this acts to minimise the space within which an alternative viewpoint could be placed. Alternatively where an utterance is framed to allow for such differing viewpoints, this actively allows and accounts for their potential placement within the dialogue. Simon-Vandenberg (2000) and Stubbs (2006) suggest that this difference between fact and opinion needs to be accounted for to assess the functionality of phrases

such as *I think*, *I believe* or *I guess*'. White (2003: 264) argues that 'the meaning or rhetorical functionality of *I think* varies according to the nature of the proposition it modulates'.

Where an utterance is dialogically expansive, Martin and White (2005) break this down into two further sub-branches:

- ENTERTAIN
- CONTRIBUTION

'ENTERTAIN' accounts for wordings which indicate that the position taken by the authorial voice is but one of a range of possibilities and therefore creates dialogic space for these other possibilities. 'When viewed dialogistically (rather than from the perspective of a truth-functional semantics, as is often the case), such locutions are seen actively to construe a heteroglossic backdrop for the text' (Martin and White 2005: 105).

'CONTRIBUTION' deals with formulations where the authorial voice disassociates itself from the evaluation and projects onto an external source. This can be achieved through the use of communicative process verbs, nominalisations of these processes and adverbial adjuncts. In some cases single lexemes such as *believe* or *suspect* could appear in both CONTRIBUTION and ENTERTAIN but the distinction is made according to whether the authorial voice takes ownership of the process or presents it as belonging to someone else. CONTRIBUTION could also be considered part of evidentiality, outlined above, in that it demonstrates 'attitudes to knowledge' (Chafe 1986) with regards to the speaker referencing the source of their information.

CONTRIBUTION is split into two further sub-branches of CONTRIBUTION: ACKNOWLEDGE and CONTRIBUTION: DISTANCE. In the former, there is the suggestion of agreement with the speaker to whom the evaluation is attributed to – *X said, X believes, according to X*. In the latter the opposite is true, with the suggestion that the authorial voice is distancing

itself from the speaker whom they are attributing the evaluation to – *X claims that, it's rumoured that.*

Even against a dialogic backdrop it is possible for an utterance to contract, rather than expand, the dialogic space. Meanings which act in this way are understood in two distinct ways:

- PROCLAMATION
- DISCLAIM

Martin and White (2005: 121) state that PROCLAMATION accounts for 'those formulations which, rather than directly rejecting or overruling a contrary position, act to limit the scope of dialogistic alternatives in the ongoing colloquy'. This is then split into two further sub-branches of PROCLAMATION: PRONOUNCE, and PROCLAMATION: ENDORSE. PROCLAMATION: PRONOUNCE deals with those situations where the authorial voice explicitly interjects itself into the utterance, *I would say, to me* or *I contend*. PROCLAMATION: ENDORSE refers to 'those propositions sourced to external sources [that] are construed by the authorial voice as correct, valid, undeniable or otherwise maximally warrantable' (Martin and White 2005: 126). This is similar to CONTRIBUTION: ACKNOWLEDGE but has the additional factor of the overt endorsement of the speaker.

DISCLAIM covers utterances which act to invoke an alternate position in order to then explicitly reject it. This is split into two further sub-branches of DISCLAIM: DENY and DISCLAIM: COUNTER. The former uses presupposition to introduce the concept of and then negate alternate positions or views. The denial may be against beliefs which the authorial voices 'assumes that at least some members of his/her mass audience will be subject to' (Martin and White 2003: 119). Tottie (1987) and Pagano (1994) have also worked in this area and coined the term 'implicit negation' to account for situations where a speaker projects a position onto the addressee for the sole purpose of then denying that position. This is commonly known as a straw man argument.

DISCLAIM: COUNTER includes utterances which represent the current position as replacing one which might have been expected. It is 'typically conveyed via conjunctions and connectives such as *although, however, yet* and *but*' (Martin and White 2003: 120). These sub-branches of ENGAGEMENT are similar to the distinctions that Conrad and Biber (2000) make under their framework of Epistemic Stance. As mentioned in section 2.2.1, these are seen as resources for not only showing the speaker's level of attachment to the proposition, but also for negotiating interpersonal relationships between themselves and other interlocutors.

In terms of this research, ENGAGEMENT will be helpful in measuring the way in which the speakers weaken or strengthen their evaluations by taking either a mono or heteroglossic approach. This distinction is obviously important, because the strength of the respondents' evaluations is likely to be a consideration when deciding how much importance is placed on the findings of the research in terms of any practical implications that it brings about. If the respondents are adamant that a certain design or design feature is extremely likeable or desirable, then this could carry more weight than if they suggested that it was possibly something that they like but that they thought other people might disagree or were not that sure or prepared to commit. Further ways in which speakers can weaken or strengthen their evaluations are introduced below.

3.2.3 Graduation

GRADUATION is concerned with values which provide a scale on which intensity can be marked, offering speakers further opportunities to strengthen or weaken their utterances. This can be done in two distinct ways, each different from the possibilities mapped out under ENGAGEMENT above. These subsections are termed FORCE and FOCUS, both of which can be used to assess the strength and direction of evaluative language. They can be distinguished accordingly:

'GRADUATION operates across two axes of scalability – that of grading according to intensity or amount [FORCE], and that of

grading according to prototypicality and the preciseness by which boundaries are drawn [FOCUS]' (Martin & White 2005: 137)

FOCUS deals with issues closely related to Prototype Theory which is discussed in detail in section 4.3.5.3. When making an evaluation, a speaker can increase its strength by stating that the target of the evaluation is absolutely typical of its kind. This would be tagged as FOCUS: SHARPEN. Alternatively the speaker can weaken the evaluation by stating that the target is only on the periphery of its type. This difference can be seen in the following statements:

*49. That **completely** that looks like one of them air fresheners.*

*50. It **sort of** looks **a bit** like an air freshener.*

In the first example the speaker has said that they think that the stimulus is as similar to an air freshener, and everything that this connotes from an evaluative perspective, as possible. In the second example the speaker has stated that is less similar and that therefore any connotations are also weaker. By stating that something is on the edge of a prototype boundary, this leaves room to then negotiate away from this evaluation at a later date. In common with ENGAGEMENT, the language that is covered under the GRADUATION branch of the Appraisal Theory Framework has previously been assessed in the literature under different terms. Softening values have been discussed in the literature, as introduced in section 2.2.3, under the headings hedges (Lakoff 1972) and vague language (Channell 1994), and sharpening values have been discussed under the headings of intensifiers, boosters and amplifiers (Labov 1984 and Hyland 2000). In addition to being applied to experiential categories, FOCUS can also be used on attitudinal categories and to decrease or strengthen other evaluative forms:

*51. I'd be **kind of** embarrassed.*

*52. That would be **sort of** embarrassing.*

*53. It to me **it's like** a young design a trendy design.*

Martin and White (2005: 138) use 'upset' to exemplify this and state that the use of such words or phrases '...construes the speaker's feelings as lying on the borderline of *upset-ness*, as having only a marginal/ non-prototypical membership in the category'. This is clearly similar to what has previously been discussed under Hedging, with Brown and Levinson (1987: 145) describing such lexical items as a 'particle, word or phrase that modifies the degree of membership of a predicate or noun phrase in a set, it says the membership is partial or true in only certain respects'. This again highlights how the use of this technique can weaken the strength of the evaluation in a way that is similar to taking a heteroglossic approach as described above under the ENGAGEMENT branch of the framework.

Where FOCUS is related to prototypicality, FORCE is instead to do with the scalability of the intensity of the evaluation. In many cases this up or down scaling is realised by an individual lexical item that has no other function such as, *a bit, somewhat, relatively, fairly, rather, very, extremely, utterly, quite* etc. Each of these can be used by a speaker to either strengthen or weaken their evaluation. Comparatives or superlatives are also included as a way of scaling intensity in a localised or relative way (Martin and White 2005). Where the up-scaling is construed as being at the highest possible intensity this is termed 'maximisation' (Quirk et al 1985). This would include locution such as *absolutely, thoroughly, completely, perfectly* and so on. In addition to these grammatical items, FORCE can also be realised lexically: *ice cold, crystal clear, dirt poor*. However these items have often been through the process of delexicalisation as there is nothing wrong, from a semantic perspective, that a drink described as *ice cold*, is likely to be just very cold. Intensification can also be realised through the use of repetition. This can either be through the repeating of an individual item or through different items from the same category:

54. *That's far far too big.*

55. *No wouldn't buy it I wouldn't buy that at all no.*

56. *That's really not very good.*

57. *I don't know I don't think that would appeal.*

Both FOCUS and FORCE allow the speaker to moderate the strength of the evaluation by controlling the specificity that is applied to the target and by manipulating the intensity. In conjunction with the approaches set out under the ENGAGEMENT branch of the Appraisal Theory Framework this covers the many different ways in which the strength of an evaluation can be both controlled and analysed. This is obviously important in relation to this research as a designer or brand owner may want to pay more attention to evaluations that have been significantly strengthened as opposed to those that have been significantly weakened.

3.2.4 Summary of Appraisal Theory

Having introduced and discussed the three key areas of Appraisal Theory, I hope to have shown that they will allow a detailed assessment of the transcripts from both the pilot study and the main study. Although each of the many sub-branches of Appraisal Theory have been introduced in this section, they will not necessarily be used in the tagging and analysis of the main study. One of the aims of the pilot study, described in section 4.2, will be to assess the suitability of the different sub-branches of the Appraisal Theory Framework. By assessing the different ways in which respondents judge and evaluate products, it might be possible to see if there are any recurrent patterns that could suggest that certain shapes, sizes or designs have a causal relationship with either positive or negative assessments. Carrying out this analysis should also help to develop a better understanding of the ways in which people make and vocalise their judgements and evaluations of products. Using this knowledge it may be possible to provide insights which could feed into the current market research beliefs and enable different insights to be achieved. This analysis and discussion will be carried out below in section 4.3. A more detailed discussion of issues related to the tagging of different evaluations in relation to their place within the framework, in addition to other issues, is carried out in section 5.4.1.1.

3.3 Summary of the Two Frameworks

Sections 3.1 provided a brief outline of Systemic Functional Grammar with the aim of providing the context from which Appraisal Theory was borne out of. I subsequently described Appraisal Theory to a much higher level of detail and due to the fact that it has been specifically tailored to help evaluate evaluative language, it is not surprising that it seems the more useful of these two frameworks. In particular the subsections under Attitude would appear to be particularly well suited to the aims and needs of this research. Although Systemic Functional Grammar could be used to help explore the evaluative language that will comprise the data for this research, it is more appropriate to use the framework that took the notions of Systemic Functional Grammar and extended them for the specific purpose of understanding evaluation in language, that is, Appraisal Theory. As such, it is Appraisal Theory that will be used to analyse the data that is collected within this research. However, the validity of any findings from this research will be reliant on more than just a solid analytical framework. The manner in which the data for this research is elicited is also extremely important and the validity of any findings will also be dependent on the design and implementation of the research methodology. This is an issue which will be discussed in Chapter 4. Having carried out a literature review of each of these areas, the next chapter will now assess the relative strengths and weaknesses of different research methodologies for measuring consumer attitudes. It will then detail the research methods that will be used in this research project before discussing the pilot study.

4. Preliminary methodology and pilot study

The three goals of this chapter are to:

- 1) Explore different data elicitation approaches and discuss the relevant issues regarding their use.
- 2) Determine the methodology for this research.
- 3) Carry out and write-up a small scale pilot study.

4.1 Issues in Interviews and Questionnaire Design

Questionnaire design and the use of interviews are likely to be at the very core of market research practices that might wish to elicit data suitable for linguistic analysis. The way in which the data is collected is obviously important in terms of both the validity and reliability of the research. This means that from the very start methodological issues in interviews and questionnaire design are going to affect the validity of the research process. It is therefore important to understand the benefits and problems that each method has in relation to eliciting, coding and analysing data. An introduction and discussion of these issues will be the focus of sections 4.1.1 and 4.1.2.

4.1.1 Interview Types

All interview types share the common goal of trying to elicit information from a respondent. 'These different kinds of interview share some common features, such as the eliciting of information by the interviewer from the interviewee' (Bryman 2001: 106). However, despite this common aim, there are many different approaches to doing this and each has benefits and problems. Each of the many different approaches and goals of research interviews fall into either of two main categories: standardised interviews or exploratory interviews.

4.1.1.1 Standardised Interviews

Standardised interviews are designed to be explicitly structured and formal. The aim is to ensure that '...every respondent has been asked the same questions, with the same meaning, in the same words, same intonation, same sequence, in the same setting and so on' (Oppenheim, 1992: 67). In this way they are comparable to questionnaires where the aim is usually to collect factual data rather than ideas and opinions. As the interviewer has far less room to move away from a pre-planned guide or script, if the guide is carefully constructed, there is less chance for them to introduce their own biases. Bryman (2001: 107) suggests that 'the standardization of both the asking and the recording of answers means that, if it is properly executed, variation in people's replies will be due to 'true' or 'real' variation and not due to the interview context. However, it should also be considered that due to the more formal nature of the interview the respondent is potentially more likely to stick to rational, 'correct' or desirable answers, as discussed below. Given the structured nature of the interviews it is imperative that the questions are worded as carefully as possible and that thought is given to the order in which they are presented. 'Each question should be worded so that the respondent understands its meaning and so that the question has the same meaning to each respondent' (Frankfort-Nachmias and Nachmias 2000: 240). Davis (1976) argues that: 'Slight changes in question wordings produce distinct effects on item distributions'.

This could be seen in Roper's study as far back as 1940. He found that support for free speech in America was twenty-one per cent higher when respondents were asked '*Do you think the United States should forbid public speeches against democracy?*' Compared to when they were asked '*Do you think the United States should allow public speeches against democracy?*' This swing could have been because of the strength of a word such as 'forbid' but regardless of the reason it shows the importance of phrasing a question in as neutral a way as possible. Once a set of neutral questions has been drawn up, it is also important to place them in an order that will have as little impact on the answers as possible. 'A major threat to the interpretation of any question form difference is the possible impact of preceding parts of the questionnaire' (Schuman and Presser 1981: 23).

The communication is also less like normal conversation so it is harder to get to the respondents' underlying beliefs or opinions. Standardised interviews are also likely to be much shorter in duration and this means that not only is it possible to carry out more of them but also that respondents are more likely to be willing to give up a shorter amount of time than the longer amount required for an in-depth interview. The standardised questions are also more likely to reduce variations in answers and this makes the data much easier to code and analyse. Despite these apparent advantages however, there are still problems with this interview type. An initial problem is that despite being methodical and precise, it is impossible to ensure that each interview is identical in structure. Bryman (2001) suggests that the characteristics of the interviewer may have an impact on the respondents' replies. Schuman and Presser (1981) showed that the ethnicity of the interviewer caused significant differences in respondents' answers when ethnicity was an inherent part of the question, for example '*Name your favourite three actors*'. It is also inevitable that there will be factors outside the researcher's control that could affect the data. External factors affecting the mood of the respondent are also likely to affect their responses and this means that the responses are inherently dynamic rather than static. This places a high import on the setting of the interview as it should be as neutral and calming as possible so that the respondent is only thinking about the questions that they are being asked rather than letting themselves be affected by external factors. Another problem is that the very nature of the interview means that the respondent is not offering their opinions on the subject matter. They are instead offering their opinions on what the researcher thinks is important about the subject matter, and this is an important difference.

Using standardised interviews could have several advantages over self report questionnaires. One of the main problems with questionnaires is the low response rate; typically only thirty per cent or so of questionnaires sent out will be returned (Silverman 2004). Having an interviewer present ensures that there will be a much higher response rate. It also means that any misunderstandings or problems can be dealt with and it will be possible to ascertain whether or not the respondent is giving genuine responses rather

than just giving any answer. All of this means that the data is more likely to be valid and reliable. The presence of the interviewer does again raise issues of interviewer bias but the structured nature of the process helps to minimise this problem.

4.1.1.2 Exploratory Interviews

Exploratory interviews are superficially similar in kind to a conversation between two or more people. They are not explicitly structured and the topic and order of question is markedly different for each interview and more free flowing.

'With little or no direction from the interviewer, respondents are encouraged to relate their experiences, to describe whatever events seem significant to them, to provide their own definitions of their situations and to reveal their opinions and attitudes as they see fit'.

(Frankfort-Nachmias and Nachmias 2000: 214)

A good exploratory interview is likely to give a high level of insight into the respondent's true feelings, thoughts and opinions, as the interview acts in a facilitating, rather than constricting, role. In comparison, the interviewer in a standardised interview can be seen as restricting the direction of the interview and therefore the answers that the respondent can give (Bryman 2001). Therefore this exploratory type of interview is most appropriate for situations where the aim is to explore the opinions and attitudes of the respondent rather than to collect specific answers to set, ordered and carefully worded questions. 'Exploratory interviews can be used to get acquainted with the phrasing and concepts used by a population of respondents.' (Fielding and Thomas 2001: 125). However, although Silverman (2004: 140) suggests that exploratory interviews are 'special forms of conversation', the flow of the communication needs to be carefully managed to ensure that the majority of the talking comes from the respondent. Every time the interviewer speaks they are likely to affect the response of the respondent in some way by introducing their own biases towards preferred responses or their own opinions on the subject matter of the interview. The use of probe questions is one area that is particularly

problematic because 'the interviewer's intervention may influence the respondent. A potential source of variability in respondents' replies that does not reflect 'true' variation is introduced' (Bryman 2001:118). The interviewer has to avoid asking leading questions or moving the focus of the interview away from where the respondent would naturally take it. 'The ideal free-style interview would consist of a continuous monologue by the respondent on the topic of the research, punctuated now and again by an 'uhuh, uhuh' from the interviewer' (Oppenheim 1992: 73). This ideal, however, is unlikely to be achieved and there are several other ways in which the interviewer's presence could affect the data. These will be discussed in section 5.4.2

Whilst it is acknowledged that there are inherent problems with this interview type, they have to be balanced against the benefits that a more standardised procedure does not provide. Although it is possible to minimise interviewer bias, create as natural an interview environment as possible and to choose the respondent sample carefully, these issues will still have an effect on the data drawn from these interviews. The main advantage of the exploratory interview is that it provides a far greater level of insight into the underlying beliefs, opinions, and thoughts of the respondent. The aim of the research is to elicit and explore natural evaluative talk so it is important that the interview methods that are used have as little negative impact as possible.

Despite the fact that there will inevitably be some level of distortion for the reasons discussed above, these interviews will still provide more naturalistic and valid data than formal standardised interviews. It was therefore decided that exploratory interviews would be used in the pilot study with a view to using them in the main study. The standard focus group or depth interview that makes up the orthodoxy within consumer insight research is also run along the lines of an exploratory interview. (Myers 1998, 2004; Smith and Fletcher 2006, and Stewart, Shamdasani, and Rook 2007). Once it was decided that this approach would be used it was then necessary to decide on the number of respondents who would be interviewed at a time. It is possible to use this approach in settings ranging from individual interviews, paired depth-interviews, triads or focus groups with up to eight respondents.

However, to look in detail at participants' linguistic output, with a view to understanding how they articulate evaluation, it is necessary to have very accurate transcriptions of what has been said by each participant. It is likely that there would be more interruptions and overlapping speech in larger groups which would also increase the difficulty of making accurate transcriptions. Therefore in addition to the practical issues of requiring longer to transcribe the interviews, it would also have made it more difficult to achieve the same level of accuracy. In larger focus groups it is also more likely that there would be more language that would simply be contesting the floor or trying to manage the topic. This in turn would lead to there being less time, and therefore language, spent evaluating the stimuli. While it would have been possible to use individual depth interviews, this would have halved the number of participants whose opinions would have been collected. For this reason it was decided that paired depth interviews would be used as this was the right compromise between being able to speak to as many respondents as was necessary and ensuring that the research setting would be best set up to elicit the type of language that is required for the transcription and analysis of the data. Paired depth interviews allow the participants to react to and build on each other's comments or observations while at the same time providing each individual with time to think about their comments while the other respondent is speaking.

Having settled on paired depth interviews to use in the pilot test it was also necessary to explore the available self-report methods to assess their benefits and problems and to trial them in the pilot study.

4.1.2 Self-Report Methods

Having looked at the different interview techniques that are available, this section will now discuss differing self-report methods that can be used to capture participant evaluations. Self-report methods can be understood as any technique where the participant is actively and knowingly providing the researcher with their data. Covert observational methods are therefore not included here as the individual is an unknowing

and non-consenting participant. In addition to the self-report interviews, as discussed above, there are also several other methodologies that have been developed to help elicit data from a participant.

4.1.2.1 Triadic sorting (Kelly Grids)

Triadic Sorting is a simplified version of Kelly's Repertory Grids (Kelly 1955). Kelly suggested that people understand the world around them through the eyes of a scientist, constantly anticipating outcomes and making hypotheses that are either accepted or rejected depending on what they observe. The subjects of the continual hypothesising might be ideas such as 'I don't like person A because...' or 'I don't get on with people who are taller than me'. The technique was initially developed out of Kelly's work in clinical psychology, trying to understand the reasons for an individual's behaviour towards other people. In terms of this research his approach can be modified so that instead of being a means for investigating a person's conceptual structure relevant to inter-personal relations, it can be used to analyse inter-object relations. In this instance the objects would be the different products. Siraj-Blatchford (1995: 195) describes it as '...a means by which the researcher may identify the fundamental categories by which a range of products or artefacts are differentiated by individuals and groups'. The respondent is shown three or more different objects and then asked to separate them into two groups on the basis of an important attribute on which the same-group products are alike and at the same time different from the other group. A respondent could be presented with a group of three bottles with the following characteristics:

Bottle A (A wine bottle) – tall, green, non-transparent, glass, holds liquid.

Bottle B (A pill bottle) – short, plastic, transparent, green, does not hold liquid.

Bottle C (A shampoo bottle) – plastic, short, non-transparent, holds liquid, blue.

These are just possible examples, as the participant would not be given a list of attributes but rather encouraged to separate them and then explain the reasons for splitting them into these differing groups.

A and B may be placed together because they are both green and C is not.

B and C may be placed together because they are both plastic and A is not.

A and C may be placed together because they are both non-transparent.

At the same time as Kelly Grids establish the attributes which are important to the respondent, they also force the respondent to use the language that they would freely use to describe the attribute. If they separate the objects by height, width, size or surface feeling, they are likely to use adjectives describing physical properties such as 'taller', 'broader', 'bigger', or 'smoother'. Alongside this they may also add an adjective which is descriptive of more abstract properties. 'It's tall and elegant'; 'these ones are smoother and more stylish', 'that one is smaller and childlike' and so on. However, Reynolds and Gutman (1988) suggest that this methodology is most suitable for concrete attributes so it may be necessary to probe the participants in more detail to elicit evaluations relating to the more abstract attributes of the pack. To this purpose, Siraj-Blatchford (1995) suggests that the respondent should be asked to provide the antonym of the adjectives that they produce and also to state which end of the scale they prefer and their reasons for this preference. If they put two products together on the basis that they are heavy, they might then give the antonym 'light'. Their preference might be for the product to be light because it is more portable. This would then elicit another important attribute – portability. This process of obtaining deeper structures is termed 'laddering' (Siraj-Blatchford 1995). In addition to this, Beck-Larsen and Nielsen (1999) suggest that these concrete attributes, such as 'tall', 'plastic' or 'heavy' are less important to the consumer than abstract attributes such as 'inspiring', 'stylish' or 'cute'. However, it is possible that certain physical, concrete properties have strong links to abstract properties. For instance 'tall' and 'thin' might be equated to 'sleek and stylish'. Certain materials, shapes or colours might equate to attributes such as 'modern', 'traditional' or 'boring'. A corpus study of consumer evaluations could be used to analyse the collocation frequency of certain words, or alternatively, participants could be presented with a list of twenty concrete, and twenty abstract properties and then asked to match them up in pairs. By triangulating the research and using both methodologies it may be possible to conclude

that there is a link between concrete and abstract attributes. This would then perhaps refute Beck-Larsen and Nielsen's claim that concrete attributes are less important. If concrete attributes act as triggers or precursors to more abstract attributes, then they could be viewed as being of equal importance. While this might demonstrate a link between abstract and concrete properties, Koller's (2008: 399) paper on the social construction of attitudes towards colour argues that 'people are culturally socialized into colour meanings. What is associated with a colour or shade is indicative not of the colour itself but of the cultural and historical formation in which it is constructed'. That is, while a link might exist in some cultures or markets, it might be different in others.

4.1.2.2 Free Sorting and Ranking

Free Sorting and Ranking are both slight variations of Triadic Sorting (Kelly 1955). Free Sorting is exactly the same as Triadic Sorting except for variations in the size of the groups. Rather than using three objects, any number can be used and there is no limit to the size of the groups that the products are split into. This means that it may be possible to introduce a more varied set of designs which would include more attributes or features than from using just three bottles. Also, if all three bottles were made of the same material, then the respondent would not be able to sort them by material so the researcher would not be able to explore whether or not material was an important attribute for that particular product. With more different designs in the group there would be less chance of all of them being the same on one particular attribute. If they were all the same on a specific attribute then it is likely that this would be for a specific reason, perhaps to do with the practicalities of the design, rather than because there was not enough variety amongst the group on display. With ranking there are no groups but instead the objects are placed in order of preference and the respondent is asked to give reasons for their choices.

4.1.2.3 Direct Elicitation

Using Direct Elicitation, the respondent is shown all of the prototypes at the same time and asked to name the most important attributes. Beck-Larsen and Nielsen (1999)

argue that this is most representative of natural speech and that it is therefore more likely to produce intrinsic, rather than extrinsic, attributes.

Having outlined some relevant issues regarding different data elicitation approaches and qualitative research interviews, section 4.2 will describe the small pilot study that was carried out.

4.2 Pilot Study

Pilot studies can be used to trial different methodologies or approaches that it is anticipated might be used in the larger scale main study. By carrying out a smaller study, it is likely that certain problems or issues will become apparent. Changes or adaptations can then be made to try to ensure that the same problems do not occur in the main study. Having reviewed analytical approaches in Chapter 3 and methodological approaches above, it was felt necessary to run this Pilot Study for two main reasons.

- 1) To check that the data elicitation techniques that had been decided on would be appropriate in terms of both the quality and quantity of language that was generated.
- 2) To assess the suitability of Appraisal Theory as the main analytical approach.

This section will explain the methodology that was used for the pilot study and discuss some of the problems and solutions that were involved in the determining the final research methodology.

4.2.1 Participants of the Pilot Study

Although I argued in section 1.1.5 that a screening or sampling of participants might be necessary for a full-scale study, this was not carried out in the pilot study. The main focus was on developing the analytical methodology so I decided that it would be

unnecessary to sample the participants for their prior knowledge of the market sector under analysis. Five participants were recruited through an advertisement that was placed in the University of Leeds Linguistics Department. The only requirements were that they were native speakers of English and aged between eighteen and twenty-five. The advertisement simply stated that people were required to help take part in a study looking at the design of consumer goods. It also informed potential respondents that there was a small cash incentive for taking part in the study. Morgan (1998) suggests that using incentives can be a legitimate method of motivating people to take part in research studies. The five respondents were made up of three females and two males (see table 7). It is argued below that the power difference between the researcher and the respondents should be kept to a minimum to ensure that the respondents are not providing what they perceive to be correct or inoffensive answers. In this context a balanced power difference could have been achieved by only recruiting other first year PhD candidates of the same age. The fact that there is a slight age gap between the researcher and the respondents could also have had some effect on the power relationships involved. However, due to the lack of availability of such respondents it was felt that the slight power differential that would be present by using younger undergraduate students would not be overly problematic.

Respondent	Age
M1	22
M2	19
F1	21
F2	21
F3	22

Table 7 Age and gender of the respondents.

4.2.2 Methodology of the Pilot Study

Each participant was interviewed separately for approximately thirty minutes and the interviews took place in the Affective Design Laboratory. This is a purpose built interview and observation room which is maintained by the Keyworth Institute and housed in the Mechanical Engineering department. The walls and furnishings are white and the design of the tables and chairs are plain. There are no windows or other decorations which could distract the respondent from the task at hand. It has been specifically designed to provide a neutral interview setting so that external influences and variables are kept to a minimum. Prior to the actual interview taking place, the respondents were offered refreshments and we had a brief informal chat to familiarise them with me, so that they felt comfortable talking to me, and also to help condition them to having their conversation recorded. They were also informed that their comments and opinions would remain anonymous, to further remove any feelings of self-consciousness about their answers, and they were given their cash incentive for taking part in the study. The interviews were recorded with both a minidisk player and an mp3 recorder and both of these were visible to the respondents.

Two different elicitation techniques were carried out within these interviews. The first stage was an open exploratory interview where the respondent was presented with a set of stimuli, in the form of seven different bottle shapes, and asked to describe them in terms of what they looked like, how they made them feel and what they reminded them of. I encouraged the respondents to handle the bottles and interact with them so that they were fully able to explore the various designs and dimensions of each different bottle. The seven stimuli are shown in Figure 9. To ensure that they had some context for their thoughts and judgements respondents were told that the bottles were prototype designs for a body moisturizer. However they were not told the name of the brand or whether it was a high or low prestige product. For this reason all of the stimuli were unbranded and the same plain white colour and neutral texture.

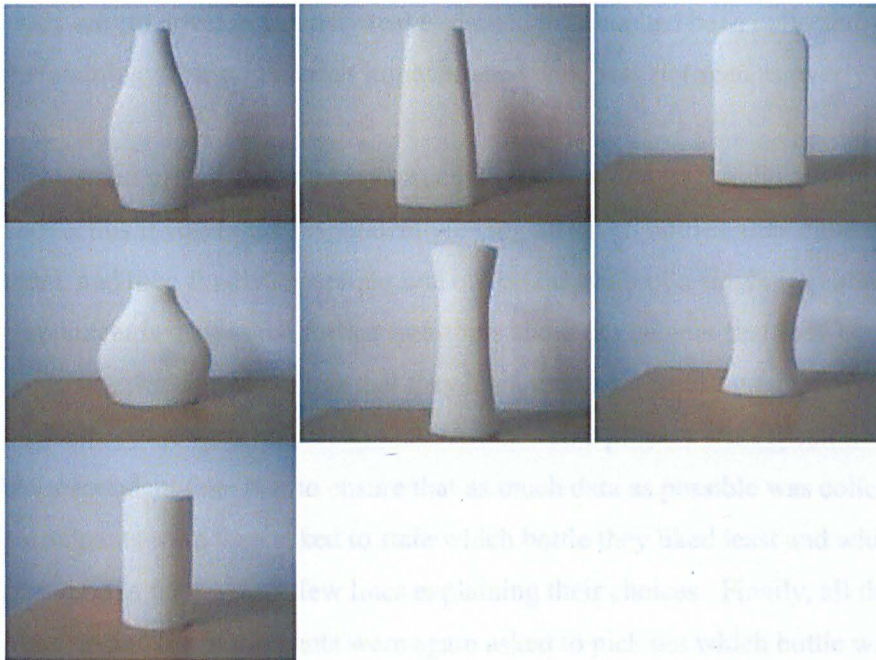


Figure 9 The seven different bottle shapes used in the experiment

As discussed above, exploratory interviews are most suitable for exploring opinions and attitudes, rather than set answers, and for discovering the phrasings and concepts that certain discourse communities use with regards to certain contexts and objects. It was therefore felt that using this technique would elicit data that would be representative of the ways in which people judge and value products. The respondents were encouraged to provide as much information as possible and I used probe questions to facilitate this. However, care was taken to ensure that this use of probe questioning caused as little leading or guiding as possible as this would have potentially invalidated the data to a certain extent. One accepted problem with this methodology is the choice of stimuli that were used. Using prototype bottle shapes to generate language about bottle shapes runs the risk of providing impoverished data in that no completely new language to do with the evaluation of bottles is likely to be elicited. Using more abstract stimuli could elicit evaluative language that would be unforthcoming from using prototype bottle shapes. However, it was felt that the respondents would require some level of context to help them make their comments and evaluations. Also, given that one of the key aims of this

study was to develop an analytical tool once the data had been collected, using a methodology that might elicit impoverished data, was not seen as overly problematic.

The second part of the interview process took the form of Triadic Sorting. As described above, this involves the respondent viewing all seven bottles, then choosing three of them, and then finally separating one out on the basis of a single important attribute. The respondent is then asked further questions about the choices that they have made to ascertain which attribute it is that they have used as separating factor and what their preferences are with regards to that attribute. This process was repeated as many times as the respondent felt able, to ensure that as much data as possible was collected. The participants were then asked to state which bottle they liked least and which they liked most and to then write a few lines explaining their choices. Finally, all the bottles were lined up and the participants were again asked to pick out which bottle was their favourite. This was done to see if there was any difference between verbal and written responses and also between reported and actual behaviour. On the completion of the interview process each of the participants was debriefed with regards to the scope of the study and the reasons for it being carried out and were again reassured that their comments would remain anonymous. Once the interviews had been completed, it was then necessary to transcribe them.

4.2.3 Transcribing the Data from the Pilot Study

Due to the fact that it was the actual content that was of primary interest for the research, a content transcription was carried out with markers for intonation, pause length and pitch were not included. Although it is suggested that 'nothing that occurs in interaction can be ruled out, a priori, as random, insignificant or irrelevant' (Atkinson and Heritage 1984: 4) it was felt that the demands that would be placed on the data would not necessitate such a detailed prosodic transcription. Fairclough (1993: 229) also proposes 'a fairly minimal type of transcription, which is adequate for many purposes. No system could conceivably show everything, and it is always a matter of Judgement, given the nature of research questions, what sort of features to show and in how much detail'.

Oliver et al (2005) define different approaches to transcription as being positioned along a continuum with two opposing modes:

1. Naturalism, in which every utterance is transcribed in as much detail as possible.
2. Denaturalism, in which idiosyncratic elements of speech (e.g., stutters, pauses, non-verbals, and involuntary vocalizations) are removed.

The transcription approach taken within this research leans more towards the denaturalism end of the scale. Although she acknowledges that some meaning might be contained within the removed content, Cameron (2001: 33) argues that a denaturalized approach can still provide a 'full and faithful transcription' and Maclean et al (2004) suggest that this is most useful for work with an interest in informational content. Oliver et al (2005: 275) argue that the 'accuracy concerns the substance of the interview, that is, the meanings and perceptions created and shared during a conversation'. It is most frequently used within various forms of ethnography (Agar 1996; Carspecken 1996), grounded theory (Charmaz 2000) and critical discourse analysis (Fairclough 1993; van Dijk 1999).

Once the transcriptions had been completed they were then ready to be analysed to assess the evaluative language that had been elicited from the participants. As discussed in section 3.4, it was decided that Appraisal Theory would be the primary theoretical approach used to analyse the evaluative language.

4.3 Analysis and Discussion

The analysis in this section will draw on each of the categories that are available under the Appraisal Theory Framework. However, a greater emphasis will be placed on the areas contained within the Attitude branch:

- AFFECT
- JUDGEMENT
- APPRECIATION

These three appear to be the most relevant in terms of the research area involved. In addition to these strands of the ATTITUDE branch, both ENGAGEMENT and GRADUATION will be used to help assess the strength of the evaluation.

4.3.1 Attitude Discussed

Martin (2003) suggests that both APPRECIATION and JUDGEMENT are very much sensitive to field and that what can be considered as Appraisal depends on the field of discourse. As such, APPRECIATION, both positive and negative, can be evoked without the use of any evaluative lexis. When ATTITUDE is realised in this manner there is inherent subjectivity in analysing both the direction of the Appraisal (positive or negative) and the FORCE (GRADUATION). There are examples of this throughout the text and they can be seen in Table 8.

Appreciation	Positive	Negative
REACTION: QUALITY	It does look like a vase as well	looks a bit like a dog's bone
	It looks a bit like sort of the old fashioned perfume bottles that you would get	it's kind of fish like in a way
	it looks like the base of a lamp	more reminds me of salt and pepper sort of shape
	it reminds me of those little perfume bottles that you get with the squeezey bit on it	that firstly reminds me of a parmesan shaker or something of that sort

Table 8 Examples of Appreciation evoked without the use of evaluative lexis.

There is no explicit evaluative lexis in these sections of the transcripts; instead the evaluation is made implicitly by comparing the bottles to objects or items that would be recognisably either positive or negative in the context of moisturiser bottles. However, while the likely intended direction of the comment can be assessed, it is still not explicitly clear as to the respondents' meaning. Therefore, to more confidently state whether they were intended as positive or negative evaluations, it is necessary to draw on other available resources. Knowledge of the field of discourse is important in ascertaining whether or not comparing the bottle to a lamp or a parmesan shaker should be viewed as a good or bad thing. The surrounding texts can also be used to create an overall impression of the speaker's opinion of the subject. Respondent one suggests that the bottle looks like an old-fashioned perfume bottle. This could be taken as a negative comment suggesting that it is out of touch or dated, or it could be a positive evaluation if old-fashioned is associated with high quality or being desirable. By looking at the surrounding text it is clear that this should be taken as a positive comment:

58. *It would look a lot nicer on your dressing table as a moisturiser bottle definitely* (REACTION: QUALITY: POSITIVE).
59. *So probably be more appealing because it looks a bit more old fashioned and that's what fashionable at the moment* (AFFECT: POSITIVE).
60. *I like that one best out of all them* (AFFECT: POSITIVE).

The same can be seen with the comparison to a 'parmesan shaker'. It is not inherently either positive or negative but the surrounding evaluative comments imply that it is meant as a criticism:

61. Speaker B: *I don't particularly like the curvature on top for some reason* (AFFECT: NEGATIVE).
Moderator: *Do you like it?*
Speaker B: *Erm no, I really don't like it actually.* (AFFECT: NEGATIVE).

The same method of drawing meaning from the surrounding clauses was also used to ascertain the direction of the other examples in table 4.

In relation to the idea of the first moment of truth, the most relevant part of the Appraisal Theory tool kit is APPRECIATION: REACTION: IMPACT. If certain design features, or combinations of features, can be shown to induce a positive response with regards to this part of the framework, then this would obviously be beneficial. An exploration of a link between the frequency of negative and positive appraisal and being the most or least liked bottle shows that there is some link, but with such a small sample it is impossible to know whether this is an indication of any kind of causality. Respondents one and four had the most number of comments that were analysed as APPRECIATION: REACTION: IMPACT: POSITIVE for the bottle that they subsequently indicated was their preferred choice. Also, respondent two had the most number of comments that were analysed as APPRECIATION: REACTION: IMPACT: NEGATIVE for the bottle that they subsequently indicated was their least preferred choice. However, with a much bigger sample, it would be possible to spot trends of this type. There is a similar link between the frequency of comments marked for positive and negative Affect and the stimulus then being identified as the one liked least or most. Again, with a larger sample size it would be possible to ascertain how strong the link was and whether or not the findings might be replicable.

Another important point to consider is the consistency of the results that are elicited. If specific features consistently produce the same responses, either positive or negative, then this would also be beneficial in allowing the producer to explore which features to include and which to avoid. With a larger sample it would also be possible to compare the results across different demographic groups such as age, gender or, as suggested in the introduction, levels of knowledge of the product under investigation.

The bottle with the highest inter-respondent consistency was bottle number two. Disregarding respondent four, for whom this was the least favourite bottle, the consensus of opinion was extremely strong. None of the other four respondents made a negative comment about it but did produce positive comments ranging from a low of four to a high

of nine. Although the distribution was slightly varied, each respondent made positive evaluations with regard to QUALITY, BALANCE, VALUATION and AFFECT. Although this is only a small sample, it does seem to suggest that this methodology may show that certain designs consistently elicit certain results and that results in certain categories affect subsequent overall levels of preference.

There did seem to be some consensus with regards to overall preferences and dislikes, with two of the respondents making exactly the same choices. Two other respondents also chose the same bottle as their least liked. However, one of the bottles that was chosen as their least liked by two respondents was chosen as the one most liked by another. As stated above, a larger sample would be required to make any broad statements but it does show the potential analytical advantages of this methodology.

4.3.2 Engagement Discussed

As stated above, ENGAGEMENT explores the way in which a speaker controls the dialogic space that their evaluation inhabits. The most obvious way of analysing this is to look at the modality of the clauses and to assess features such as hedging and mitigation. Table 4 seems to show some kind of pattern in terms of the level of engagement that respondents made with their comments. In these instances, it seems that respondents had a stronger level of ENGAGEMENT when making positive statements than when making negative statements. Each of the negative statements is mitigated or hedged, for example *'it's kind of'*, *'looks a bit like'*, *'or something of that sort'* and *'sort of shape'*. In comparison, only one of the positive evaluations has these features *'it looks bit like sort of...'* and the rest are definite statements *'it does look like'*, *'it looks like...'* and *'it reminds me of...'*. This pattern is repeated throughout the transcripts with the respondents making negative evaluations of the products, but in doing so, softening the strength of their comments:

62. *I don't know if it would attract that much attention* (REACTION: IMPACT: NEGATIVE) *I don't know.*
63. *Don't know so that makes it look less striking.* (REACTION: IMPACT: NEGATIVE)
64. *It's probably a bit tall for how thin it is* (COMPOSITION: BALANCE: NEGATIVE) *I don't know though.*

In each of these examples, the respondent makes a negative appraisal but the use of '*I don't know*' seems to act as a disclaimer. In contrast, the same respondent, commenting on the same stimuli produced the following comments:

65. *I like* (AFFECT: POSITIVE) *the design of it.*
66. *I like* (AFFECT: POSITIVE) *the design of this one better.*
67. *That would be nice* (REACTION: QUALITY: POSITIVE) *as a cosmetics bottle.*

These examples of positive appraisal are all unmitigated and do not have the same softening features that occurred in the previous examples of negative appraisal.

The use of low value modals in conjunction with negative appraisal is also prevalent within the transcripts:

68. *This might seem a little more design for design's sake* (BALANCE: COMPLEXITY: NEGATIVE).
69. *It's kind of fish like in a way a fish without a head or tail* (REACTION: QUALITY: NEGATIVE) *perhaps.*
70. *I think maybe because it's been done so much before* (VALUATION: NEGATIVE) *perhaps.*

In contrast there are very few examples of positive appraisal combined with modals of a low value. Indeed all but one of the high value modals occur when the respondent is making a positive appraisal of the product:

71. *It would look a lot nicer on your dressing table as a moisturiser bottle*
(REACTION: QUALITY: POSITIVE) *definitely*.
72. *I definitely like* (AFFECT: POSITIVE) *that one the best*.
73. *I'd definitely buy* (VALUATION: POSITIVE) *this shorter one*.

One explanation for this could be that despite the best efforts of the interview, it is possible that there was a perceived power difference between the respondent and the interviewer and the respondents may have felt reluctant to make strong, unmitigated criticisms. Fairclough (1993: 159) suggests that 'low affinity with a proposition may express lack of power, rather than lack of conviction or knowledge'.

Brown and Levinson's (1987) notion of face may also have been a factor in this discrepancy between the level of ENGAGEMENT for positive and negative appraisal. If the respondents viewed the bottle shapes as belonging to the interviewer then they may have felt that to criticise them would be a face threatening act. The idea that there was an overall preference for the respondents to praise, rather than criticise, is supported by the comparative frequencies of positive and negative evaluation. In total there were 106 occurrences of positive evaluation and just 59 negative evaluations.

Another reason for the more cautious approach that the respondents took towards critically appraising the products could be the very nature of an experimental setting. In normal, everyday conversations, opinions are quite short lived and not normally called to account in a detailed manner. The fact that the respondents knew that their answers were being recorded, and would be analysed in some way, may have made them more cautious in expressing opinions or made them feel that they should give what they perceived to be 'correct' answers. Although every effort was made to help the respondents feel at ease, using a non-naturalistic research methodology does have this disadvantage.

4.3.3 Graduation Discussed

As discussed in section 3.3.3, GRADUATION is concerned with the scaling of intensity of a respondent's comments. With regards to the most obvious demonstrations of FORCE, adverbial intensifiers, the split between negative and positive FORCE is very similar. There were 51 FORCE markers of this kind and of these, 27 were positive:

- 74. *It looks a **bit** like sort of the old fashioned perfume bottles that you would get.*
- 75. *So yea I like that it's that's **slightly** different I think.*
- 76. *I **really** like that one I think.*

The remaining 24 were negative:

- 77. *A **bit** too normal.*
- 78. *An element of **completely** unnecessary to design a bottle in that fashion.*
- 79. *I don't **really** like this one.*

Table eight shows that the split between the different levels of FORCE was also similar. The distribution of FORCE markers was also equal amongst the different stimuli; none of the bottles produced an unexpectedly higher or lower frequency of evaluative markers of this kind. However, it is interesting to note that there was nearly double the frequency of low intensity markers of FORCE than high intensity markers. This again suggests that the respondents may have been less happy about making strong, explicit opinions that could be called into question.

	Positive Force	Negative Force
Bit	16	14
Slightly	3	0
Completely	0	2
Very	3	3
Really	5	5

Table 6 A table to show the split between Positive and Negative Force.

As suggested as above, there is some overlap between FORCE and ENGAGEMENT, as both are related to the level of conviction that the speaker has about their comments. Markers of FORCE such as *kind of*, *slightly* and *a bit* have a low intensity and are comparable to the hedging and mitigating discussed in section 2.2.3. High intensity markers of FORCE such as *completely*, *very* and *really* are similar to the higher levels of ENGAGEMENT, also discussed above.

The scaling of intensity can also be lowered through ‘a process of narrowing or broadening the terms by which category membership is determined, through the sharpening or softening of semantic focus’ (www.grammatics.com 2005). In the instances that occur within the transcripts, the intensity is always lowered. An opinion is stated ‘*it’s narrow*’ but instead of definitely being narrow the descriptive category is blurred and becomes ‘*it’s sort of narrow*’. This acts to mitigate the speaker’s opinion because they have not made such a specific claim about the stimulus. While the split between positive and negative FORCE was broadly equal, Table 9 shows that the split between positive and negative FOCUS is unequal.

	Positive FOCUS	Negative FOCUS
Kind of	8	4
Sort of	20	6

Table 7 The split between Positive and Negative FOCUS

This difference could be caused by the fact that while the respondents were fairly sure of whether or not they liked a bottle, they were less sure as to the reasons why, and so tried to keep the FOCUS of the evaluation much broader. Another resource for controlling the strength of an evaluative comment is the potential to make the opinion either INSCRIBED or EVOKED.

4.3.4 Inscribed and Evoked Appraisal

Martin (2003) describes INVOKED APPRAISAL as being explicitly expressed within the text as opposed to EVOKED APPRAISAL which is projected by reference to events or states of being which are conventionally prized. An explicit opinion leaves little room for misunderstanding within the intended meaning of the comment *'makes me think a bit bland'* (REACTION: IMPACT: NEGATIVE). This can safely be regarded as a negative comment as there is no obvious situation where a moisturiser bottle would want to be considered bland. Another example from the same respondent, evaluating the same stimulus, is less instantly obvious in its criticism *'looks a bit like a dog's bone'* (REACTION: QUALITY: NEGATIVE). Comparing a moisturiser bottle to a bone is not inherently critical; the respondent could then go on to state that there were qualities of a dog's bone which they held in high regard, and that for them it was a positive comparison. However, in the absence of such comments, most readings of such a statement would suggest that it was a negative evaluation. As discussed in section 4.3.1, this can be further justified by assessing the overall evaluation of the respondent's feelings towards the stimulus. In this instance there were five negative evaluations and three positive ones, none of which were explicit. Respondent three's least liked bottle had a higher proportion of explicit criticism than any of the other bottles and their preferred choice also contained the most examples of explicit praise with three out of seven comments:

80. *I definitely like (AFFECT: POSITIVE: EXPLICIT) that one the best.*

81. *I just think it works nicely (VALUATION: POSITIVE: EXPLICIT).*

82. *It's like yea that's quite cool (VALUATION: POSITIVE: EXPLICIT).*

This pattern is followed with respondent number five. Their favourite bottle had the second most comments and again it provoked more explicit comments than the other stimuli without having any negative comments at all:

83. *So yea I like (AFFECT: POSITIVE: EXPLICIT) that it's slightly different I think.*

84. *But I quite like it its different erm yea very simple (COMPOSITION: BALANCE: POSITIVE: EXPLICIT) but not boring.*

None of the bottles which were picked out as being the respondent's favourite had any negative comments made about them. In contrast three of the five respondents still gave some positive comments with regards to the bottle they liked least.

The other two respondents also made more comments about their favourite bottles than ones that they felt less strongly about. The stimuli that the respondents liked least also provoked a higher number of comments which could suggest that where they had strong opinions about an object, they had more to say regardless of evaluative orientation. The fact that there were more explicit comments made with regards to the most and least liked bottles also suggests that when the respondents' opinions were stronger, they were more prepared to make them clearer. Fielding and Thomas (2001) suggest that respondents may tend towards less polarised answers if they feel that the power difference is too large or the setting too formal. Therefore, if the respondents felt that they had good reason to give a stronger opinion, they might have felt that they would be better able to support and justify that opinion if they were questioned further on it. If they were less certain about their judgement then they might not have wanted to provide a strong opinion which they would feel less sure about being able to defend. Of the bottles which provoked the most evaluations, the majority of them were either liked least or most by their respondent. Having analysed the data from the exploratory interviews, the same framework will now be applied to the data elicited from the Triadic Sorting section of the Pilot Study.

4.3.5 Triadic Sorting Analysed

Both Triadic Sorting and Exploratory Interviews were used within the pilot study to elicit as much data as possible from the respondents. The aim of using these two different approaches was to assess if they elicit particularly different data, and if so, to explore which, if either, methodology is most appropriate for the research aims of this study. This section will compare the data that were elicited by these two different techniques, and also discuss some of the issues regarding the data that were elicited using this approach.

4.3.5.1 Points of Comparison

This section will see whether the findings that were discussed in section 4.3 were replicated when using this alternate methodology. Many of the observations that were made in section 4.3 were in relation to a specific stimulus, so the aim was to explore three different areas:

- 1) Were the same types of evaluative language consistently used?
- 2) Was the stimulus consistently (dis)liked?
- 3) Did the overall preference of the bottle affect the quantity of language elicited?

With exploratory interviews this was possible because each stimulus was looked at in turn. However the Triadic Sorting approach does not do this and it is therefore not possible to make these kinds of assessments. The data elicited from the Triadic Sorting can be used to verify some of the key findings from the exploratory interviews. Two observations which are closely related are that evaluative language is not always explicit and not always in the form of single word items. This can be seen in the following examples:

85. *It gives me that feeling of Hispanic womanliness* (REACTION: IMPACT: POSITIVE).
86. *Ok erm these two reminds me of two perfumes I've got at home* (REACTION: IMPACT: POSITIVE).
87. *I think I associate it with something so like the figure of the woman* (REACTION: IMPACT: POSITIVE).

The comparative nature of evaluation is obviously highlighted by this methodology as the participants are actively encouraged to compare the stimulus, so it is not surprising that there is a higher level of comparative evaluation. However, in addition to inter-stimulus comparisons, there are also comparisons to internal ideas or prototypes:

88. *I wouldn't have classed that as a like bottle shape so maybe that is a bit more unusual* (REACTION: IMPACT: POSITIVE).
89. *Erm I just prefer the curved the shapes to the straight lines* (COMPOSITION: BALANCE POSITIVE).
90. *Maybe that one cos it's smaller and fits into the size of your hand erm whereas these ones are bigger and a lot less stable* (COMPOSITION: BALANCE: NEGATIVE).

Although some of the findings can be verified, there are others that cannot be and this, as well as other problems with the data elicited from the Triadic Sorting, are discussed below.

4.3.5.2 Problems with Triadic Sorting

The most obvious initial point of comparison is that much less data were elicited from the Triadic Sorting. One of the reasons for this could be that with each participant the triadic sorting was carried out after the exploratory interview. This means that they could have been fatigued or bored by the time they were asked to carry out the Triadic Sorting and therefore found it much more difficult to provide an opinion. One simple way to overcome this problem would be to alternate the order in which the interviews were carried out. In addition to providing less data, this methodology also produced data of a lower quality in terms of its use for this research. To successfully apply Appraisal Theory, there needs to be as much free-flowing language as possible. The nature of Triadic Sorting requires the interviewer to take a more active role in managing the conversation and asking questions. Part of this greater level of activity was due to the Triadic Sorting process being more complicated. Therefore one way of reducing this involvement would be to explain what was required in more detail either at the start of the whole interview process or just prior to the Triadic Sorting stage. Instructions were provided on the work sheet (see Appendix 2) but there was still some level of confusion on the participants' behalf when it came to carrying out the task.

The Triadic Sorting also tended to produce much more descriptive, rather than explicitly evaluative language. Although it was suggested in section 2.4.3 that it is possible to assess the direction of implicit evaluations by drawing on the surrounding language and evaluations, this is not always possible. Much of the language elicited with this technique was deictic in nature.

91. Moderator: *So which do you prefer out of those two?*

Speaker A: *These two groups?*

Moderator: *Yea.*

Speaker A: *Erm well I don't mind I like **this** group because it's got **that** one in it.*

Moderator: *Ok.*

Speaker A: *But I don't like it's got **that** one in it but I like **that** group erm is it more what I would buy?*

Moderator: *Yea well yea which you prefer.*

Speaker A: *Which I prefer generally well I'll say **that** one then.*

This means that it is not always possible to ascertain which bottles or group the respondent preferred and consequently it is not possible to assess whether the previous evaluation was meant positively or negatively.

If this approach is to be used for the main study then there are several ways in which these problems can be avoided. Firstly it might be necessary to have a video recording of the interviews as well as the audio recording. This would then solve the problems associated with the use of deictic language as it would be possible to see what 'I prefer that one' was actually referring to. Simply being aware of the problem also means that I would be better able to avoid it by managing the interviews more efficiently in terms of the language that was elicited. However there would then be the issue of leading the participant to produce data that matched up to what was wanted or required.

This approach highlights the issue of whether or not people actually view products in their own right, or whether or not they make use of comparative methods to help with their evaluations. One theory that can help with understanding this is Prototype Theory.

4.3.5.3 Prototype Theory

Taylor (1995) suggests that there are two different ways in which the human mind assesses how typical of its genre an object is. There is the Aristotelian principle which suggests that there are ‘...necessary and sufficient conditions for membership, with clear cut boundaries, and with only two degrees of membership, i.e. member and non-member’ (Taylor 1995: 59). The opposing way in which objects are categorised is explained by Prototype Theory. This proposes that everything we see fits more or less into ready made internal categories.

‘A prototype category, loosely defined, is a category with a clear core or central members of the category, but variables or even fuzzy boundaries. The core members have a cluster of properties, but the peripheral members of the category lack some of the core properties... psychological experiments indicate that these prototypes do play a role in categorisation by human beings’
(Croft 2003:162).

There are criticisms of Prototype Theory with regards to some of its deeper interpretations (Goodluck 1991; Lee 2002) but the notion that there are categories to which we ascribe objects is fairly well established (Langacker, 1987; Field 2003; Janda 2006). Understanding how we categorise objects in relation to other items is important as it helps to understand how we view them individually. Of particular interest is whether or not consumers look at objects in isolation, or whether they do something different. The buying environment for most products is clearly a comparative one, with rival products placed alongside each other. If consumers usually make judgements in a

comparative way then this knowledge should be incorporated into any research method seeking to better understand and influence consumer decision making.

Given the nature of the research methods used in this study, triadic sorting in particular, it would be expected that there would be many examples of comparative evaluative language. However, the responses to the first bottle that each respondent was shown could indicate whether or not they were comparing what they were shown to a prototypical moisturiser bottle. The very first comments of two of the respondents seem to suggest that they may have had a prototype image in their mind of what a moisturiser bottle should look like. By comparing the shape to that of other cosmetics products they may have been indicating that they felt that the shape that they were presented with was on the periphery of the category 'moisturiser bottle'.

- 92. *Erm it looks like sort of a something you'd find in a shower gel bottle.*
- 93. *Erm, it reminds me of a shampoo bottle.*

The other respondents also made comparative statements in response to being shown the first bottle:

- 94. *It's just something you'd normally see.*
- 95. *You know it's not a typical shape.*
- 96. *Wouldn't really see it as a moisturiser bottle.*

There was no suggestion that comparisons should be made and there was nothing physical for them to make comparisons with, so this could also be an indication that they were comparing the stimuli in front of them to a mental image of a prototypical moisturiser bottle. There were also many instances of more explicit comparisons between bottles:

- 97. *That's like the erm the other one it's just a lot shorter and wider.*
- 98. *I like the design of this one better than the other one.*

99. A bit thicker this one a bit more chunky.

Together with the use of similes, this appears to demonstrate that when viewing an object, even in absolute or relative isolation, internal images are used to help form evaluative opinions and judgements:

100. Like an old fashioned talcum powder.

101. Looks a bit like a dog's bone.

102. It looks like the base of a lamp.

This might suggest that consumers already have an idea of what they expect their products to look and can call upon a 'standard' image, for comparison, when looking at new designs.

4.4 Summary of the Pilot Study

The aims of the pilot study were to carry out, analyse and discuss five research interviews with the purpose of testing the suitability of both the elicitation and analytical approaches that were used. With regards to the elicitation techniques, it is apparent that the language that was generated from the exploratory interviews was definitely suitable for analysing with Appraisal Theory. Although not a naturalistic approach, the language that was generated was free flowing and conversational in style. Due to careful management it was possible to ensure that the participants held the conversational floor for the majority of the time with the moderator limited to occasional probing and topic management where necessary. The participants were able to speak freely about each of the stimuli for as long as they felt able and the use of a discussion guide ensured that all of the necessary areas were sufficiently covered.

As discussed above, elements of the Triadic Sorting stage of the interviews were less successful. Due to the more physical, rather than verbal, nature of the task, much of the

language that was elicited was deictic in nature. There was also much less language generated and the moderator was required to have a far higher level of input due to the more complicated nature of the task. This higher involvement increased the risk of influencing the participants and also contributed to the lower level of language that was generated. Although it may have been possible to overcome some of these problems through the use of video recording, more detailed instructions, and an awareness of the need to encourage more fully constructed language, it was decided that these would not necessarily be practical and would add substantially to the work load. For these reasons it was decided that Triadic Sorting would not be used as substantially in the final research methodology. Part of the reason for using more than one data elicitation methodology was to make it possible to triangulate the data. Once it was decided that Triadic Sorting would not be used, it was then necessary to use a different approach. The Ranking technique was used in the Pilot Study and this was found to be quick and easy to do. Although it did not generate any language, it did mean that it was possible to compare the evaluations that were made in the exploratory interviews with explicitly given evaluations in terms of preference levels. Therefore it was decided that the main data elicitation techniques that would be used for the main study would be Exploratory Interviews followed by a simple Ranking exercise. However, because the final research project was carried out in conjunction with Unilever and their Consumer Insight Team, they had some input into the final data collection methodology.

Other questions which were raised as a result of the Pilot Study were:

- 1) Is showing the stimuli in isolation an appropriate approach, bearing in mind that the buying environment is comparative and people may be using internal categories to compare the stimuli to an existing prototype?
- 2) Would it be beneficial to split the sample by the respondents' prior use and knowledge of the market area under investigation, as those with strong feelings about the products may elicit a higher frequency of more useful data?

However, it is important to remember that this was a very small-scale study and that further research, on a much larger scale, needs to be carried out to assess whether any findings made here can be replicated and validated. With a larger sample it would also be possible to compare findings across different demographic groups such as age, gender and location on the Register General's Social Scale, and across different levels of knowledge of the market under investigation.

In terms of the analytical approach that was used, it was quickly apparent that Appraisal Theory would be an extremely useful tool to use. Although not designed for a market research purpose, many of the categories map directly onto the kind of areas that researchers might be interested in. Where there were areas that did not map directly onto the language that was generated, it should be possible to account for these evaluations with only small adaptations to the framework. However, having analysed the data it was also apparent that it would not be necessary to use the entire framework to the fullest of levels as it would not add any significant knowledge for the purpose of market research and new product development. Hood (2004: 74) argues that:

'As with any system network within SFL, a principle of delicacy applies, so that movement from left to right across the networks represents more general to more specific analysis. Any analysis of data can select an appropriate level of delicacy for coding.'

In particular, the ENGAGEMENT and AFFECT branches are likely to be simplified. While this might miss out some subtleties in terms of the way in which the hetero- or monoglossic stance is framed, it is broadly in line with Martin and White's (2005) thinking.

5. Final methodology and main study

Having overviewed work on evaluative language, explored different linguistic analytical approaches, discussed methodological issues in data elicitation, and carried out a small pilot study, this section will explain the methodological approaches of the main study on which this research is based.

5.1 Introducing the Main Study

The aim of this chapter is to outline and justify the data collection methods that were used in this research. Due to this research being an industry sponsored case studentship, some of the decisions were made in conjunction with the placement company. This is something that will be discussed below. The data elicitation process that was developed in conjunction with Unilever was comprised of four stages and each of these will be discussed in the following sections. These four stages were:

- Initial Focus Groups.
- CAD Sessions.
- Reassessing the Design.
- Paired Depth Interviews.

As part of the study, it was also decided that the stimuli would be developed as part of the process, rather than simply presenting them to the participants. This meant that there were two sets of participants; those taking part in the development of the stimuli and those taking part in the final data elicitation interviews.

5.1.1 The Participants

As will be discussed below, thirty hours of data were collected for this research, to ensure that there was a sufficient quantity of evaluative language to analyse. Paired

depth-interviews were used in the final stage and this meant that sixty people would be required for this stage. From the experience of the pilot study it was expected that the participants would be able to talk about each stimulus for approximately eight minutes. In addition to talking about each bottle in isolation there were also other stages within the interview process (see section 5.2.4.1). With each interview lasting one hour, this meant that there would be time to discuss five prototype packs within each interview. The process for developing the 5 stimuli is detailed below from 5.2 onwards.

5.1.1.1 The Unilever Panel

Due to the large amount of consumer research that they carry out, Unilever maintain a panel of consumer volunteers who take part in their research. The participants are screened to ensure that they are not members of other industry panels and to ensure that they do not take part in consumer insight work on too frequent a basis. Each member of the panel is restricted to two sessions a year. The participants drawn from the panel for this research were aged between eighteen and thirty-five at the request of Unilever. The only other restriction was that they were regular users of deodorants, antiperspirants or body sprays. The recruitment letter simply stated that they would be taking part in interviews about deodorant packs and that it would be looking at their shape and design, rather than being concerned with the contents. This research formed part of an existing Unilever project being run for the Lynx brand of deodorants meaning that Unilever wanted the stimuli to be developed from a male consumer point of view. This meant that the first focus group was comprised of six male members of the Unilever panel. However, once this group was completed, it quickly became apparent that it might be necessary to redo this initial group with different participants. As shown in Figures eleven to fourteen, the sketches that were made by the Unilever panel members were all very similar to current packs on the market and they were lacking in innovation and creativity. The sketches are all very similar, both to each other and to current deodorant packs in the market. The main points of similarity are their general size, size of the base, ability to stand up, and the point of release of the contents. The participants for the paired-depth interviews, however, were still drawn from the Unilever panel as there was

no requirement for these participants to have any specific knowledge or abilities with regards to packaging design. Although Unilever were particularly interested in developing a packaging design for the male consumer, they also wanted to involve female respondents as from their consumer tracking data they know that it is often the female in the household who buys male personal care category products. Therefore both male and female members of their consumer panel were used in the paired depth interviews.

5.1.1.2 Skilled Participants

Having reviewed the initial group it was felt that the participants needed to have some understanding of products and design to be able to carry out the tasks required in the initial focus group session. One of the aims of the project that this research was part of was to explore new and innovative designs. In conjunction with Unilever, I therefore decided to redo this group with participants who had a demonstrable interest and ability on packaging and design. These participants were recruited by emailing students on relevant courses at the University of Leeds. The aims, set-up and results of these sessions will be introduced and discussed in section 5.2.

5.2 Methodology of the Main Study: A Four Stage Process

As stated in the introduction to this chapter, four different stages were developed to create the stimuli and elicit the data for this study:

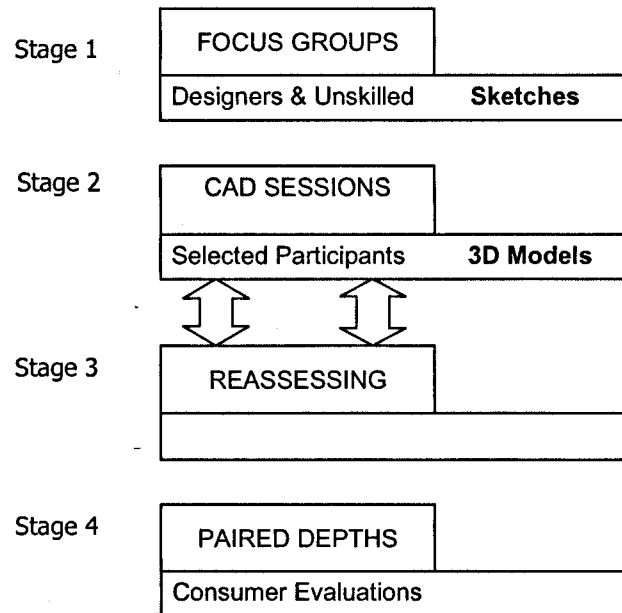


Figure 10 Four Stages of Development

Figure 10 shows the way in which the process worked through these four stages:

- Firstly, there were initial focus groups to explore consumers' general opinions on current deodorants and their packaging. These groups also generated participant sketches of their ideal packs.
- Secondly, these sketches were modelled using CAD software as part of an individual interview with each selected participant.
- Thirdly there would be an opportunity for the participants to reassess their designs once they had seen a 3D model of it and to change it until it completely matched their initial concept. Stages two and three could be repeated as many times as was felt necessary to ensure the prototype was exactly as the participant had intended.

- Finally there were the paired depth-interviews which were carried out with new participants and used the five prototype designs as stimuli. It was from these final interviews that the data was collected.

5.2.1 Stage 1: Initial Focus Groups

The aim of these sessions was two-fold. There were the industrial aims of exploring general consumer perceptions of deodorants and looking at the impact that different packaging solutions have on brands and product perceptions. There was also the research aim of obtaining sketches from the participants of their ideal packaging design. The initial group was comprised of six male members of the Unilever Panel. In following with standard practice, the session started with the participants introducing themselves to each other and being made aware of the recording equipment in the room and the viewing window at the back of the room.

As discussed in section 4.1.1.2, exploratory interviews are useful when the aim is to explore the opinions and attitudes of the respondent rather than to collect specific answers to set, ordered and carefully worded questions. As such, although a discussion guide was written and used for these initial groups (Appendix 1), the aim was to allow the flow of conversation to be managed by the participants whilst keeping in mind certain topics that needed to be covered. To encourage the participants to think about different packaging types and possibilities, they were shown a wide variety of different deodorant packs that were currently on the market. These were then discussed with regards to the effect it had on their perceptions of the product. The main areas that were discussed were:

1. Reasons why they used their current deodorant.
2. What they expected from their deodorant pack.
3. The impact of packaging and design on their purchasing decisions.

4. The idea of borrowing design features from products from other categories.

In addition to these stimuli, concept boards from previous Unilever projects were also displayed to the participants. These boards contained images of possible new designs and packs for a deodorant. Once these had been discussed, the participants were then asked to consider all the topics that had been discussed in the duration of the session and sketch out a design that they felt was the ideal packaging for the deodorant that they were currently using.



Figure 11 Sketch 1



Figure 12 Sketch 2



Figure 13 Sketch 3



Figure 14 Sketch 4



Figure 15 Sketch 5

Once these initial sessions had been completed there were sixteen sketches to choose from. As stated above, it was felt that five stimuli would be required so it was then necessary to decide which sketches would be used. This decision was made by Unilever as it was felt that all of the sketches would be suitable for the research requirements and

there were some designs that they were more interested in exploring, than others. The chosen sketches can be seen below in Figures sixteen to twenty.

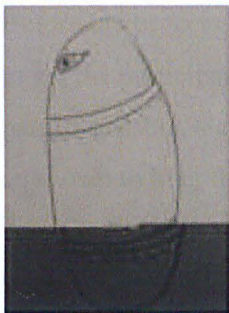


Figure 16 Design A

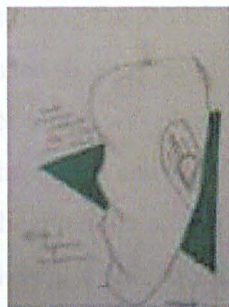


Figure 17 Design B

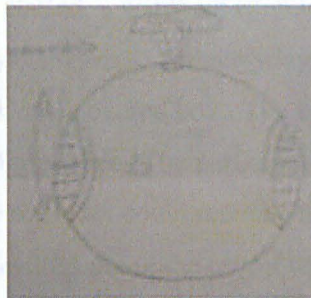


Figure 18 Design C

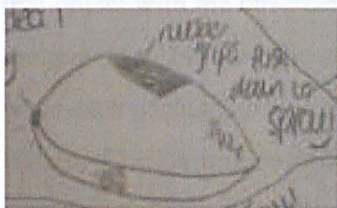


Figure 19 Design D



Figure 20 Design E

Once the sketches had been chosen it was then necessary to build them into real life prototypes. This was done with the aid of a CAD modelling process. CAD software enables a modeller to create a 3D image on screen that can be adjusted and tailored in as much detail as required. This image can then be 'printed' to make a real 3D object. The aims, set-up and results of these sessions will be introduced and discussed in section 5.2.2.

5.2.2 Stage 2: The CAD Sessions

The aim of the CAD sessions was to end up with five 3D images that fairly represented the sketches made by the participants. Once the sketches and their designer had been selected, the participant was brought back in for a session with a CAD operator and me. The aim of these sessions was focused on a practical output, creating the model,

so although there was a general discussion guide (see Appendix II) the structure of these sessions was less controlled.

Before the process of creating the model was started, a warm-up exercise was carried out to get the participants used to using descriptive language. An everyday object such as a pair of scissors, a cup or a fork was placed inside a covered box. The object was unknown to both the participant and the CAD modeller. The participant was asked to describe the object to the modeller so that the modeller could recreate the object on screen. The modeller was also able to ask questions about the object with regards to its dimensions and shape. Once both parties were happy with the onscreen image, the object was removed from the box and both the participant and modeller could then compare the real object with the onscreen version. This process was carried out twice for each participant. This also allowed the CAD operator to become accustomed to the different ways each participant made their descriptions. The main purpose of the session was then started. The starting point was a very rough outline shape based on the original participant sketch. To help with the scale of the model, a pencil was accurately modelled prior to the session (see figure 21).

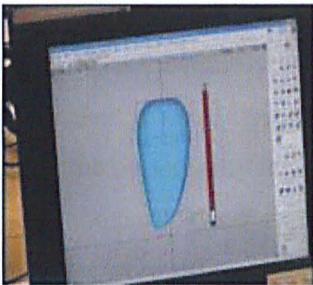


Figure 21 Still From CAD Session

The pencil was also placed on the desk for the participant to refer to when working out the desired dimensions for the CAD model. Although the participants were questioned as to the reasons behind their choices, the main aim was to end up with a modelled packaging design.

I felt that it was important to bring the participant into this stage of the prototype development because otherwise the CAD modeller would have been interpreting the

initial sketch unaided and important details may have been lost. Once the participant was happy that the model fully resembled their initial idea and sketch, the CAD operator prepared the model to be created.

5.2.3 Stage 3: Reassessing the Prototype

Although the CAD sessions were only finished when both the CAD operator and the participant were happy with the outcome, it was still felt necessary to allow the participants to have an opportunity to reassess the design once it had been created and they could feel it in their hands. It was anticipated that there might be features or ideas that whilst looking good on paper, would be viewed differently with a real prototype.

Ideally this session would be run along the same lines as stage two but with the addition of the prototype in the participant's hands. This would allow them to make any adjustments that they felt were required. Again, when both participant and modeller were happy with the final result, it would be possible to create a real model of the prototype. It would then be possible to repeat stages two and three as many times as was feasible to ensure that the prototype was absolutely perfect. Due to the time restraints of the placement it was only possible to show the participants their prototypes without offering them the opportunity to make any changes. However, in these reassessment interviews, all of the participants expressed a desire to make alterations to their designs. Some of the changes were simply in terms of the overall size of the prototype but some of there were more detailed. With design B (see figure 17) it became apparent that in addition to the overall size of the design being too big, the grips were also impractical and wrongly positioned which meant that it could only be used by people who were right-handed. Once this process had been completed for each of the five chosen prototype designs, it was then possible to explore consumer perceptions of them. Although these first three sections were recorded, it was with the expectation that they might be used to add weight to findings, based on the analysis of the main interviews, rather than being analysed in their entirety themselves.

5.2.4 Stage 4: Paired Depth-Interviews

Once the previous three stages had been completed, it was possible to carry out the paired depth-interviews. The aim of these interviews was to elicit evaluative language from the participants, using the five prototypes as the main stimuli. Thirty hours' worth of data was seen as a necessary compromise between the desire to collect as much data as possible and the need to keep the transcription and analysis time requirements to a manageable level.

Each session started with the participants introducing themselves to each other and being made aware of the recording equipment in the room and the viewing window at the back of the room. They were then told that they would be shown five different shapes that were designs of packs for a deodorant. It was made clear that each stimulus would be the same in terms of colour and texture and that they would all be unbranded. This was done to ensure that the evaluative comments were based on the shape of the stimuli and not any other variables. As in the recruitment letter, I told the participants that I was interested in their opinions on the shapes and designs. As discussed in section 2.1, perceived power differences between the interviewer and respondent can affect the responses. I therefore made it clear to the participants that I did not work for Unilever and that I had played no part in the design of the shapes. This was done to try and minimise the power difference between myself the participants and to try and ensure that they would be comfortable criticising the stimuli rather than giving positive responses for reasons of politeness. The first stimulus was then brought out. In most cases, the participants spontaneously started to make evaluative judgements about it without any prompting from myself. In these cases it was then simply a task of maintaining the conversation by using probe questions and ensuring that both participants were providing their opinions. It was also necessary to ensure that all of the pre-chosen topics were covered and that there was enough time to cover all of the tasks in the session. To aid

this, a discussion guide was drawn up (see appendix III). The key areas that were explored were:

- What are your opinions of this pack, do you (dis)like it?
- Who do you think this pack is aimed at?
- Do you see this as more of a cheap design or a premium design?
- How do you think this pack would look on the shelf?

Further, more specific questions were then asked, depending on the initial responses, to probe the participants' opinions further. This process was repeated for each of the stimuli. In each interview the ordering of the stimuli was randomised to minimize the effects of repetition, fatigue or other order effects.

When moderating, it often seemed that the sections of the interviews about the packs that were most liked were the most productive and easiest to carry out, but this is not borne out by the facts. Although it was intended that each prototype would be discussed for broadly the same period of time, no direct action was taken to control this. The participants were encouraged to talk about the stimuli for as long as they were able to produce new comments or opinions but if they had little to say about a particular pack then although the discussion guide was followed, there was no minimum time to be aimed for. Though there was a clock in the room, this was only checked periodically to ensure that the whole session would not overrun the allotted time and was not used to monitor each section. As discussed in the introduction to chapter 6, the data for this study is drawn from the sections of the interviews covering the most and least liked stimulus based on the final ranking exercise, as introduced below. In total, the average time spent talking about these two stimulus was very similar. For the most liked the average time was 8 minutes and 4 seconds and for the least liked it was 8 minutes and 12 seconds. However it is worth pointing out that for the majority of the interviews the participants actually spent marginally longer discussing least liked, but there were also 2 interviews where the participants spent nearly double the length of time talking about the most liked and this brought the average times closer together. This raises the question of

whether people have more to say when they feel strongly about something, regardless of polarity, or whether they say more about things that they like. This issue will be explored in more detail in Chapter 7. Once the general discussion about each of the stimulus was completed, the second stage of the interview was carried out.

5.2.4.1 Triadic Sorting

As discussed in section 4.1.2.1, Triadic Sorting can be used to elicit further information from participants that might not be forthcoming in an exploratory interview. It is also useful in helping to determine the attributes that an object is comprised of and by using a laddering technique it is possible to explore consumers' opinions on these different attributes. The aim of this part of the interview was to elicit further data from the participants. However, as noted in section 4.4, there were significant problems with this technique in terms of the language that it generated both in terms of the quality and the quantity. It was therefore decided that this approach would play a less significant role in the main study. Despite this it was still felt that it was worthwhile retaining the process as part of the methodology as the Unilever team was interested in seeing how participants responded to it as a data elicitation methodology. All of the pack designs were placed on the table and the participants were asked to take it in turns to choose two packs that had something in common that they either liked or disliked. They were then asked to choose a pack that contrasted the pair they had just made. Once this was done I asked them to explain what the shared feature was of the pair and why it was important to them. This was done as many times as possible with the participants being encouraged to have as many goes as possible and to think about all of the concepts and areas that had been discussed in the main interview. Once this had been finished, the final stage of the interview was carried out.

5.2.4.2 Ranking

As discussed in section 4.1.2.2, Ranking can be used to directly ascertain the order of preference of different objects. At the end of the session the participants were

asked to rank the packs in order of preference and to comment on why each pack had been placed where it had. This was done to help triangulate the findings from the Appraisal Theory analysis. If participants give the most positive, and least negative, comments regarding the pack which they subsequently ranked highest, then this would help to validate this approach to consumer insight. Knowing how each participant has explicitly ranked each pack will also be helpful when assessing the polarity of evaluative tokens which are not inherently clear.

The participants were then asked if they had any further comments on any of the designs, before the interview was finished. They were then debriefed with regards to the focus and purpose of the research.

5.3 Transcribing the Data from the Main Study

With all the data having been collected, it was then necessary to transcribe it in preparation for analysis. As suggested in section 4.2.3, the way in which the data will be used will not necessitate the inclusion of markers for intonation, pause length or pitch. Although the transcriptions will not be detailed in this manner, these prosodic features might still be used to help analyse some of the data. In some instances the polarity of an evaluative comment might be unclear. In those instances, the tone or pitch of the comment will be assessed to see if this can help determine the direction of fit.

5.4 Tagging the Data from the Main Study

Once all of the data was transcribed to the appropriate level, it was then necessary to start tagging the data for each of the parts of the Appraisal Theory Framework, its role in the turn-taking structure and in terms of who made the utterance. The following section will use examples drawn from transcripts from the main study to explain why certain evaluation types were placed within specific branches of the Appraisal Theory

Framework. Although there is inevitably the potential for a reader to take a different position (Martin and Rose 2003), Hood (2003: 113) argues that ‘a given text naturalises a particular reading position’ and this should allow the required level of consistency to be achieved. The notion of tagging evaluations in relation to the structures within which they are produced is introduced and discussed in detail in section 2.1. Section 5.4 will briefly explain how this scale is used to tag evaluation in this regard and also introduce the Participant layer which simply identifies made each individual utterance. As stated throughout this thesis, the categories of Appraisal Theory map onto many of the areas that market researchers might be interested in when measuring consumer responses to products, brands and marketing communication. Therefore many of the decisions will be self-explanatory but where they are not they will be discussed in more detail. Finally, this section will detail the use of inter-rater reliability testing that was carried out to try and maximise the validity of this research and any findings associated with it.

The data for this research was annotated and analysed using the software CorpusTool v1.14 (O'Donnell 2007). This software is designed to allow the researcher to annotate and query a corpus of linguistic data in several different ways. However, before the data could be tagged and then queried, one or more schemas needed to be developed. The schemas take the form of multiple layers within a framework with increasing levels of detail. They also have the capacity to include multiple features within a layer and this will be discussed in more detail below. Each of these schemas can also be customized in the CorpusTool software to allow the researcher to create a framework that matches their specific needs. The corpus is made up of data drawn from many different transcripts and the software enables the researcher to use the same annotation schemes across each of these different texts. When discussing Layers of Analysis, the term ‘Layers’ is used differently from the standard meaning as understood within corpus linguistics and as introduced above. Instead of referring to the different levels of detail within an annotation schema, it refers to completely separate schemas that can be used in conjunction with each other on the same data. For this research I will be using three layers to annotate and analyse the data:

- APPRAISAL
- Turn-type
- Participant

Within the Appraisal Framework, a layer is a category such as ATTITUDE which has features such as AFFECT, JUDGEMENT or APPRECIATION. Each of these features can then have further layers – in the case of APPRECIATION these are REACTION, COMPOSITION and VALUATION. Appraisal Theory's level of depth means that there are several different starting points each reaching five or six levels of complexity. The following screen shot shows how the Appraisal Theory Framework is represented in the CorpusTool software: Each evaluative unit is assessed in terms of which part of the framework it matches up to. This process runs from left to right across the different levels of the framework. A single evaluative phrase might consist of evaluative units that belong to each of the initial three APPRAISAL categories; ATTITUDE, ENGAGEMENT or GRADUATION:

103. I don't think [ENGAGEMENT] it's really [FORCE] bad [ATTITUDE].
104. That there is a bit [FOCUS] too big [ATTITUDE] to me [ENGAGEMENT].

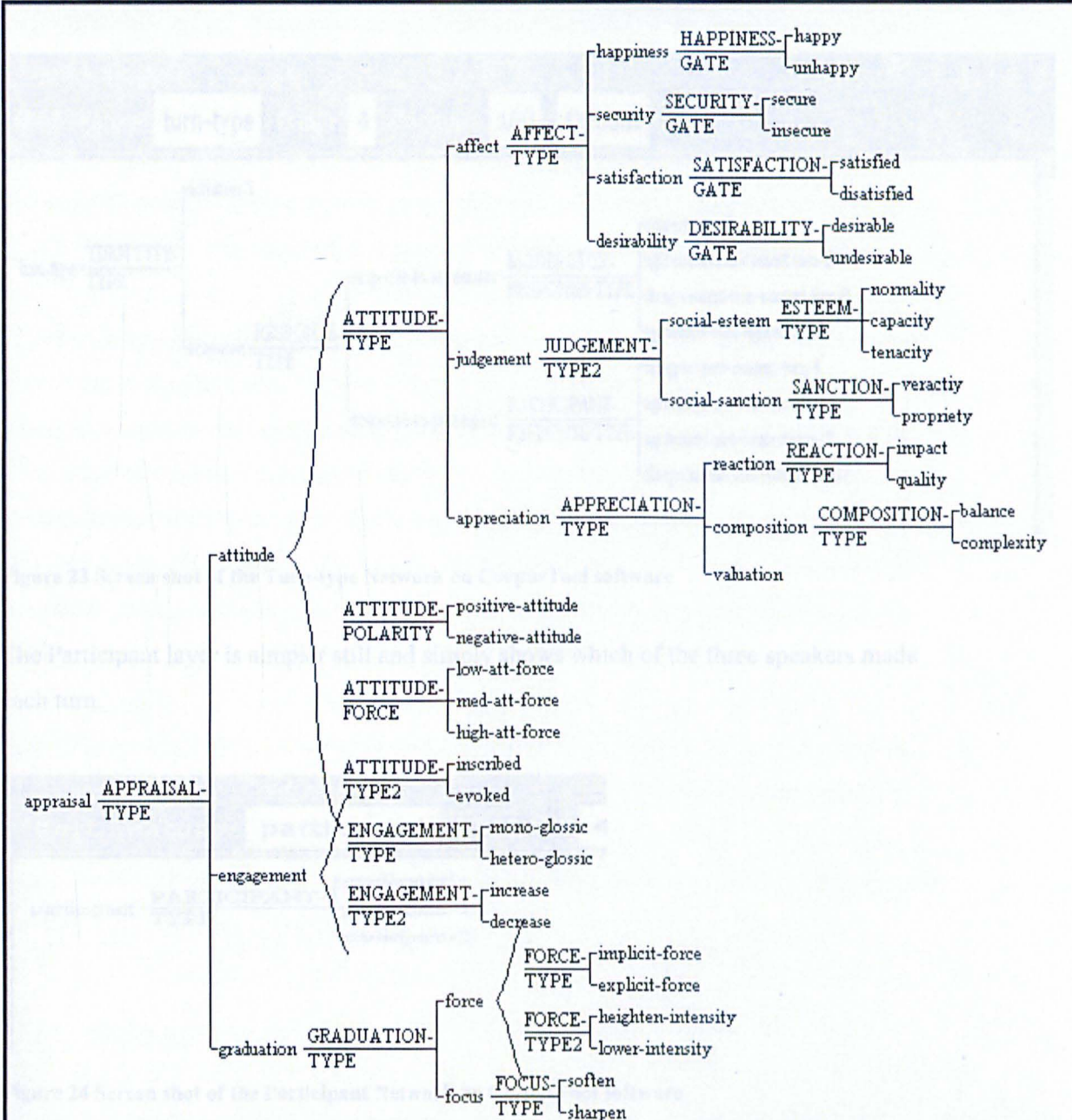


Figure 22 Screen shot of the Appraisal Network on CorpusTool software

The Turn-type layer has less depth as there are fewer potential outcomes. Here the initial layer shows whether the turn is an initiation or a response with subsequent features accounting for whom it was responding to and what kind of response it was.

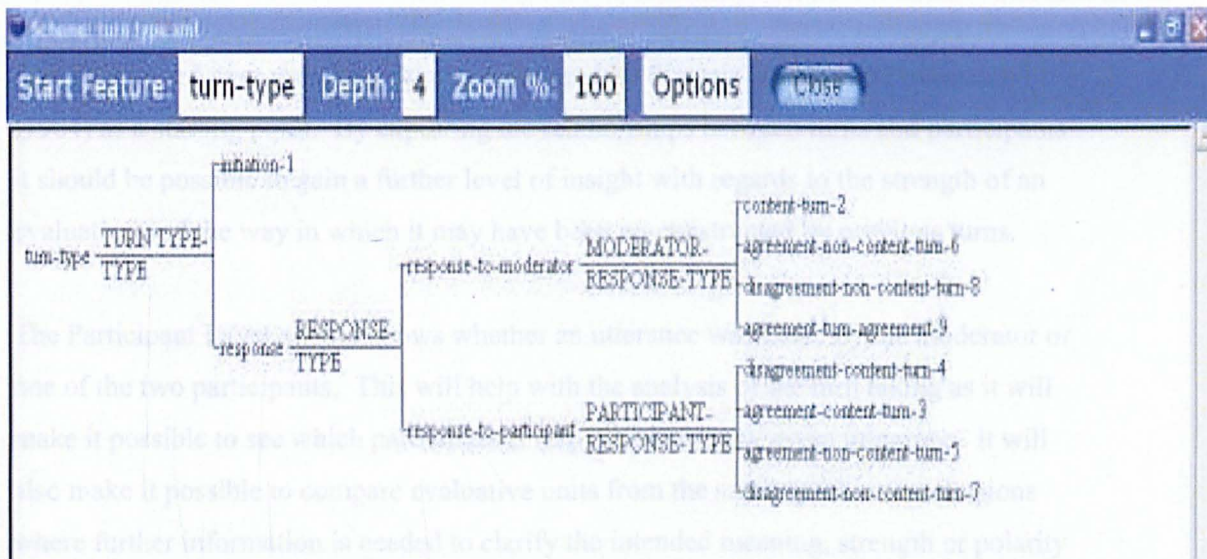


Figure 23 Screen shot of the Turn-type Network on CorpusTool software

The Participant layer is simpler still and simply shows which of the three speakers made each turn.

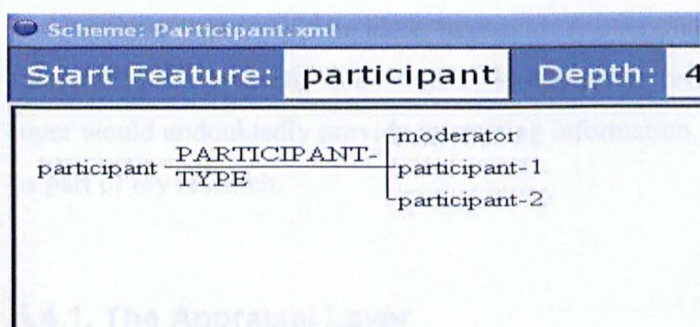


Figure 24 Screen shot of the Participant Network on CorpusTool software

This means that each unit of data will be tagged for each of the three levels. It will have a tag for the part of the Appraisal framework that it belongs to, a tag for which kind of turn-type it is and one to show which participant made the utterance. One of the main

features of the CorpusTool is that the corpus can be queried across these different 'Layers'. The Appraisal Layer is used to assign each evaluative unit to a part of the framework, which then facilitates the search for significant patterns or trends with regards to their usage. This is the deepest and most detailed of the three layers.

The Turn-type Layer uses frameworks developed by Sinclair (1975) and Pomerantz (1984) as a starting point. By exploring the relationships between turns and participants it should be possible to gain a further level of insight with regards to the strength of an evaluation and the way in which it may have been co-constructed by previous turns.

The Participant Layer simply shows whether an utterance was made by the moderator or one of the two participants. This will help with the analysis of the turn taking as it will make it possible to see which participant is responsible for any given utterance. It will also make it possible to compare evaluative units from the same speaker on occasions where further information is needed to clarify the intended meaning, strength or polarity of a less obvious evaluation. It might also be necessary to add a further Layer to account for other conversation analysis features such as tag questions, politeness markers or topic initiations.

The addition of a further layer to account for demographic markers such as age, gender and purchaser type would be likely to provide further, more detailed information of the kind that would be useful in an industry based market research setting. While such a layer would undoubtedly provide interesting information, these areas will not be explored as part of my research.

5.4.1. The Appraisal Layer

This section will explain why certain evaluations were placed within specific areas of the Appraisal Theory Framework. As noted above, many parts of the framework map directly onto the kinds of areas that would be of interest to market researchers and also onto the kind of evaluative utterances which are likely to be elicited in this context.

Therefore many of the decisions with regards to which part of the framework an evaluation should be placed were self-explanatory and unproblematic. This section will exemplify these instances but it is the occasions where it was less obvious as to which part of the framework an evaluation should be placed in that will be discussed in more detail. This section will start off by discussing tagging issues relating to the ATTITUDE branch before exploring similar issues in both ENGAGEMENT and GRADUATION.

5.4.1.1 Tagging Attitude

- As introduced in section 3.3.1, the ATTITUDE branch of the framework is specifically tailored to help account for:

- i. How a speaker is emotionally disposed to the subject of the communication.
- ii. How the subject of the communication compares to accepted norms and values.
- iii. How the subject of the communication creates an impact on the speaker in terms of form, appearance and aesthetics.

It should be apparent that these are appropriate to the kind of evaluations that are made in this research. These three sub-categories are labelled AFFECT, JUDGEMENT and APPRECIATION and are looked at in further detail below.

5.4.1.1.1 Tagging Affect

As noted in section 3.3.1.1, AFFECT is a resource for measuring emotional responses to the stimuli. Although the whole extended framework for AFFECT was introduced in section 3.3.1.1, this was simplified for the main project. Where an evaluation was tagged for AFFECT it was simply tagged as being either positive or negative and then for one of the sub-branches of HAPPINESS, SECURITY, SATISFACTION or DESIRABILITY. Determining the polarity of the evaluation was straightforward in most

instances and where it was not, it was possible to use the surrounding utterances and the broader context to help assess the intended meaning as discussed above. The main issue that occurred when tagging for AFFECT was double tagging. For some of the evaluations that were tagged for AFFECT it was simply a reporting of the emotional response '*I really like that*', '*I wouldn't want something like that*' or '*that would be embarrassing*'. In these instances they would simply be tagged for AFFECT. However, on other occasions there was a specific target for the emotional response and in these instances the same evaluation was tagged for both the AFFECT and for the focus of the evaluation:

105. *I like the hand grips.*

106. *I don't like that it's far too big.*

107. *I'm slightly concerned that the spray mechanism's exposed.*

In each of these cases there is an explicit target of the emotion and therefore it would seem to miss some of the point of the evaluation to not account for it. The above examples were therefore also tagged for COMPOSITION: COMPLEXITY, COMPOSITION and BALANCE and COMPOSITION: COMPLEXITY respectively. The idea of 'double tagging' has been discussed extensively within the Appraisal online discussion group with particular regard to evoked Affect and the way in which this can be realised. Page (2003: 216) suggests that double-tagging is appropriate where 'the appraised item may in turn evoke yet another classification at a secondary level'. She goes on to argue that given the fundamental connection between all three categories, 'In theory it would seem possible for a great many instances of APPRECIATION and JUDGEMENT to also evoke AFFECT' (Page 2003: 216). The importance that the flexibility that double-coding offers is also highlighted by Macken-Horarik (2003: 314). 'Present work in Appraisal is crucial because it explores overlaps in choices [and] possibilities of multiple coding'.

5.4.1.1.2 Tagging Judgement

As noted in the introduction to section 3.3.1.2, JUDGEMENT is primarily concerned with human behaviours and how they relate to social norms. However within the context of this research it seemed that there were some evaluations which would fit

within the JUDGEMENT: NORMALITY branch of the framework. Although there is some overlap between JUDGEMENT: NORMALITY and VALUATION there were some evaluations which seemed to best answer the trigger questions of 'How Special?' for JUDGEMENT: NORMALITY rather than 'Was it worthwhile?' for VALUATION.

108. *It's an odd shape I'nt it?*

109. *It is something out of the ordinary isn't it?*

JUDGEMENT in its original use relates to the behaviour or circumstance of a person but in the context, and for the sub-branch that I am using, it would be the normality of the object. This would be in relation to the participant's expectations of what a typical deodorant pack might look like. Evaluations tagged under VALUATION would instead relate to '...things, whether concrete or abstract, material or semiotic' (Martin and White 2005: 59). Therefore there would be different codings for 'odd' in the examples below:

110. *It's an odd pack.* JUDGEMENT: NORMALITY

111. *It's an odd way to hold it.* APPRECIATION: COMPOSITION: COMPLEXITY

So while it was possible to adapt the intended usage of the framework for the new context of market research in new product development, this was the only part of the JUDGEMENT branch that was used. With regards to the polarity of the evaluations it was usually obvious whether the participant meant it to be a positive or negative evaluation.

112. *It's a bizarre shape for a deodorant isn't it?*

113. *It's an odd shape I'nt it?*

In each of these instances it is hard to imagine a scenario where these descriptions could be viewed as positive. However there were some evaluations where it was necessary to look to the surrounding text to see whether the evaluation was intended to praise or criticize the stimulus. Each of the evaluations below could be either positive or negative dependant on whether or not the speaker liked or disliked the pack.

114. *It's not that different really.*
115. *I haven't seen anything like that before.*
116. *I don't think I've ever seen one that's laid flat.*

Looking at the surrounding evaluations can help to determine the speaker's general feelings about the pack. Looking at where the participant placed the pack in the ranking exercise was also used here.

5.4.1.1.3 Tagging Appreciation

As stated above, APPRECIATION is related to the impact that a product or object has on an individual and as such this branch of the framework maps onto the areas that are likely to be of interest to market researchers, particularly with regards to pack design and new product development. APPRECIATION is split into three sub-branches of COMPOSITION, REACTION and VALUATION with REACTION and COMPOSITION both split into two further sub-branches. These sub-branches are termed COMPOSITION: BALANCE, COMPOSITION: COMPLEXITY, REACTION: QUALITY and REACTION: IMPACT. As first described in table 3 in section 3.3.1.3, the different branches and sub-branches each have trigger questions to assess where any given evaluation should fit. The basic trigger question for COMPOSITION: BALANCE is '*Did it hang together?*' In this context, this related to how the packaging looked and how well proportioned it was physically. Therefore, where a speaker's evaluations contained comments regarding the size, physical usability or actual balance, this was tagged for COMPOSITION: BALANCE.

117. *You'd need two hands to spray it.*
118. *It's not even comfy I wouldn't have said.*
119. *It's just too big for a deodorant bottle.*

One of the key issues with regards to the different stimuli was whether or not they stood up and this was also tagged under this branch of the framework.

120. *It sort of doesn't stand up or anything.*
121. *Wouldn't sit on the shelf very well.*
122. *It's not something that's going to stand up.*

The trigger question for COMPOSITION: COMPLEXITY is '*was it hard to follow?*' This accounted for evaluations relating to how difficult the participants found the pack to use, in addition to whether they thought it was too complicated or nice and simple. Therefore everything to do with the ease of use of the pack, from a non-size related perspective, was tagged for COMPOSITION: COMPLEXITY.

123. *How would that work what's that?*
124. *It's to tell you where to put your hand.*
125. *Leads you to pick it up correctly.*

There is obviously some overlap in evaluations where the difficulty with using it was caused by the physical dimensions of the pack. In these cases, if the evaluation simply commented on the size, it was tagged as COMPOSITION: BALANCE if it commented on the size being the cause of the difficulty of use, it was tagged for both and if the difficulty was not size related it was just tagged as COMPOSITION: COMPLEXITY.

The trigger question for REACTION: IMPACT is '*Did it grab me?*'. One of the topics in the discussion guide related to how the pack would look on the shelf and if it would stand out from other pack designs already on the market. Any evaluations relating to the level of impact the pack would have were tagged for REACTION: IMPACT.

126. *You could not even notice that.*
127. *Yea I think it would attract me.*
128. *That's a bit dull.*

There is some potential overlap with regards to both the JUDGEMENT and VALUATION branches of the framework where the impact is caused by the pack being very different. The examples given by Martin (2000) for VALUATION include *Innovative, Unique* and *Insignificant* and the examples for JUDGEMENT: NORMALITY includes *Odd, Normal* and *Every day*. Where the evaluation was related to the impact the pack would have, it was tagged under this branch of the framework even if it was the normality, or otherwise, of the pack that caused the impact.

For REACTION: QUALITY the trigger question is *'Did I like it?'* This would seem to be very similar to the AFFECT branch of the framework which is actually exemplified with the verb 'like' in Martin (2003). Martin and White (2005: 57) state that:

'Clearly there are strong links between REACTION and AFFECT, including derivationally related lexis. Nevertheless we think it is important to distinguish between construing the emotions someone feels (AFFECT) and ascribing the power to trigger such feelings to things.'

Though the actual lexis that is used might be similar or even the same, the grammatical form that it takes, in addition the context within which it is used, justify the tagging across different branches. As Eggins and Slade write:

'The interpretation of the meaning of lexical items is not only dependent on the co-text but also on the sociocultural background and positioning of the interactants. Appraisal analysis must therefore be sensitive to the potential for different readings or 'hearings' of attitudinal meanings' (2001: 126).

The fact that REACTION: QUALITY is dealing with the desirability of the object also means that it again overlaps with the DESIRABILITY sub-branch of the AFFECT system. Where this was explicitly indicated through the use of an emotional marker, the evaluation was double tagged to account for both the APPRECIATION and AFFECT that is being displayed.

129. *I wouldn't really be attracted to that at all.*
130. *I really like the femininity of it.*

Another of the topics in the discussion guide was related to whether or not the participants thought that each stimulus was more likely to be aimed at a male or female market. Although evaluations relating to femininity or masculinity could perhaps be tagged under the COMPOSITION branch of the framework, I felt that the REACTION: QUALITY branch would be a more suitable fit. Therefore the following evaluations were each tagged for REACTION: QUALITY:

131. *I just think it's a very feminine shape.*
132. *I'd see that as a female one.*
133. *It's just for a man maybe but not for a woman.*

In instances such as this it is apparent that an evaluation has been made but it is less clear as to the intended polarity. Looking at the linguistic and non-linguistic context can help to determine this. This is discussed in more detail in Chapter 6. In addition to evaluations related to concepts of masculinity and femininity, some evaluations using similes were also tagged under REACTION: QUALITY.

134. *Looks like a mouse a computer mouse.*
135. *It does look like a you know a flower receptacle.*
136. *Like erm a trendy candle or something.*

A further topic in the discussion guide was whether the participants felt that the stimulus looked like it would be for a cheap or expensive deodorant brand. Evaluations relating to the perceived cost were also tagged under the REACTION: QUALITY branch as although this only indirectly links to the trigger question of 'did I like it?', issues of price seem to relate to the perceived QUALITY of the pack. Therefore the following evaluations were also tagged for REACTION: QUALITY:

137. *I think that probably is a little bit more expensive.*
138. *Certainly wouldn't think it were cheap.*
139. *Yea I'd say I'd say more expensive even though it's smaller.*

As with the polarity of evaluations related to the femininity or masculinity of the pack, the surrounding context is likely to be useful in determining whether the evaluation was intended to be positive or negative. Again this is discussed further in Chapter 6.

Examples of less problematic evaluations that were tagged for REACTION: QUALITY are shown below:

140. *Mmm it's a bit ugly isn't it?*
141. *It looks nice.*
142. *It's more decorative.*

The trigger question for VALUATION is '*Was it worthwhile?*' Where other parts of the APPRECIATION framework deal with specific parts or effects of the pack, VALUATION deals with the overall merit of it. Therefore evaluations that summarised other points that had been made or gave an overview of the participant's opinion were tagged for VALUATION. Given that the ultimate purpose of the design is to persuade consumers to buy it, evaluations relating to purchasing decisions were also tagged in this way:

143. *I think it's the best probably out of the three.*
144. *I reckon that would be a seller.*
145. *Probably I'd probably buy it cos it's a novelty.*

As these evaluations often acted as a summary of the participants' opinions the polarity was normally clear.

5.4.1.1.4 Tagging Engagement

As noted in section 3.3.2, ENGAGEMENT is related to the way in which the speaker frames their evaluation in ‘various ways [which] construe for the text a heteroglossic backdrop of prior utterances, alternative viewpoints and anticipated responses’ (Martin and White 2005: 97). Where the speaker acknowledges that there might be alternative views and explicitly accounts for that fact in their evaluation, this is a heteroglossic evaluation. Where they wish to set out their opinion as being the only possible one and act to head off any alternate views, this is a monoglossic evaluation. Tagging for ENGAGEMENT was more straightforward than for the APPRECIATION branch of the framework as many of the markers come from closed word classes such as modal auxiliary verbs:

- 146. *Might look at it and think oh that's new or that's different.*
- 147. *I suppose more masculine.*
- 148. *I'd still probably use that.*

Other mitigators or intensifiers were also prevalent throughout the evaluations, as for the reasons discussed above in section 2.1.2, the participants were likely to be continuously trying to control the perceived strength of their evaluations. Therefore all utterances which acted to permit or confront alternate viewpoints were also tagged under the ENGAGEMENT branch:

- 149. *I think it's a novelty more than anything.*
- 150. *That's definitely a very feminine shape.*
- 151. *I don't know it just doesn't appeal to me really.*

As stated above, the ENGAGEMENT branch of the framework was simplified to only account for whether the speaker had weakened, using a heteroglossic form, or strengthened, using a monoglossic form, their evaluation. Therefore the polarity of the ENGAGEMENT was always clear. While this simplification means that some of the nuances of measuring the positioning of intersubjectivity are lost, as stated in section

3.3.2, White (2003) draws a clear distinction at this higher level of the framework between those evaluations that either expand or contract the range of alternate viewpoints.

'At the broadest level, then, I make a two-way distinction between the monoglossic utterance (the undialogized bare assertion) and the heteroglossic or dialogistic utterance in which some engagement with alternative position and/or voice is signaled' (White 2003: 265).

Within my analysis, the only area where there was some debate, was with regards to what Martin and White (2005) term ENGAGEMENT: PROCLAMATION: PRONOUNCE. This is where the speaker explicitly presents themselves within the text with phrases such as 'to me' 'I would say' or 'personally'. They argue that such 'overt intervention into the text by the authorial voice... [implies] the presence of some resistance [while] challenging or heading off a particular dialogistic alternative.' They go on to argue that this shows the speaker '...interpolating himself explicitly into the text in order to indicate his maximal investment in the current proposition' (Martin & White 2005: 129). In contrast, I would argue that by framing the evaluations as belonging so explicitly to themselves, the speaker is acknowledging that other people might have a different view and that is therefore a heteroglossic and weakened form. I would argue that the acknowledgment of other positions or views shows a less than absolute certainty of a proposition, White suggest that this '...conveys a heightened personal investment in the viewpoint being advanced and thereby explicitly indicates an interest in advancing that view point.' (White 2003: 269). This seems to acknowledge that the speaker is aware of other viewpoints, but rather than accepting the potential disagreement, they argue that this tactic opposes any alternate view in the strongest possible manner. However within the same article, White (2003: 264) describes a speaker's utterance of '*I firmly believe*' as '...grounding his proposition in his own individual, contingent subjecthood and thereby representing his value position as but one among a potential diversity of viewpoints'.

5.4.1.1.5 Tagging Graduation

As discussed in section 3.3.3, GRADUATION is related to the way in which a speaker can strengthen or weaken the intensity of their evaluations. The GRADUATION branch is split into two sub-branches: FOCUS and FORCE. FOCUS deals with the way in which the specificity of subject of an evaluation can be sharpened or softened, these are the terms by which its polarity is measured, FOCUS: SHARPEN and FOCUS: SOFTEN. By honing in on exactly what it is that is being evaluated, the speaker increases the FOCUS and therefore the strength of the evaluation. Conversely, by muddying the waters a little surrounding the object of their evaluation, the speaker is allowing themselves some leeway in terms of how strongly they are evaluating a specific object, idea or event. Therefore words or phrases which increased the focus of an evaluation, and therefore its strength, were tagged FOCUS: SHARPEN:

152. *It's just the **actual** dimensions I think are too bulky to use.*
153. *It's like a **real** brick.*

Words or phrases which acted to soften the preciseness of the focus, and therefore the strength, of the evaluation were tagged FOCUS: SOFTEN:

154. *It's **like** a value pack.*
155. *I'd probably say **about** middle.*
156. *Cos it is quite **sort of** funky funky design.*

FORCE covers the way in which the intensity of an evaluation can be heightened or weakened by the speaker. Where the intensity is heightened, this is seen to strengthen the evaluation and where the intensity is lowered this is seen to weaken the evaluation.

Examples of FORCE: INTENSITY: INCREASE are:

157. *It's **very** wide though isn't it?*
158. *It is **really** thick at the top.*

159. *It's something **totally** different.*

Examples of FORCE: INTENSITY: LOWER are:

160. *It's **a bit** big and bulky.*

161. *You look as if you're getting **quite** a lot there.*

162. *I'm **slightly** concerned that spray mechanism's exposed.*

In most cases the tagging for GRADUATION was straightforward as there are clearly defined boundaries between the different categories. However there were certain cases where the polarity or category was less clear. 'More', 'Quite' and 'Rather' could each be tagged in different ways depending on their context and their prosody.

- More

163. *'Cos of the shape I think it does look **more** trendy that's what I mean by the Armani and the Bosses type.'*

The 'more' here could be evaluated as either GRADUATION: FOCUS: SOFTEN or GRADUATION: FORCE: HEIGHTEN INTENSITY. 'It's more trendy' could be taken as 'it's not completely trendy but it's on the way to being trendy'; in this way it acts as a synonym to other GRADUATION: FOCUS: SOFTEN markers such as *mostly, kind of or effectively*.

Alternatively, *It's more trendy* could be taken as 'it's not just trendy but it's more trendy', especially if it is being used as a comparative, comparing the stimulus to a specific, different object, rather than to the prototypical 'trendy' object. In this way it would be acting as a synonym to other GRADUATION: FORCE: HEIGHTEN INTENSITY markers such as *completely, rather, or very*.

- Quite

164. *'I quite like the er the groove thing.'*

Here the *'quite'* could be viewed as showing a lack of FORCE in terms of how much the speaker likes the 'groove thing'. In this way it is acting as a synonym of other GRADUATION: FORCE: LOWER INTENSITY markers such as 'I like that **a bit**' or 'I like that **a little**'. However, with a different intonation it could have the opposite effect and act to increase their level of FORCE.

- Rather

165. *'That's rather nice.'*

As with *'Quite'*, discussed above, the *'rather'* could be viewed as showing either an increase or a decrease dependent on the tone of voice and other prosodic features. In these cases the original recordings were checked to listen to the intonation. The surrounding text was again also used to help determine the most likely intended meaning.

5.4.2 The Turn-taking Layer

In Chapter 2, I discussed the literature on turn taking, preference and power, content and non-content turns and politeness theory in relation to evaluative talk. I then proposed a scale of strength of evaluations based on both the structural and social elements that are involved in the production of any given utterance. In summary I argue that:

- Initiation turns are indicative of greater evaluative force than response turns,
- Responses to the moderator are indicative of greater evaluative force than responses to a fellow respondent
- Dispreferred or disagreement turns are indicative of greater evaluative force than preferred or agreement turns,

- Content turns are indicative of greater evaluative force than non content turns

This scale is recreated here with a brief summary of why each layer is ranked in such an order.

1. Unsolicited Evaluation: Initiation Turn

This type of evaluation is considered to have the greatest inherent strength as it is made spontaneously rather than as a result of expected structural norms and is also made without the influence of leading questions or preference structures. In this way it can be considered a more 'clean' evaluation.

2. Response to Moderator: Non-content Disagreement Turn

Giving a dispreferred response to a speaker with more power carries a high potential cost to the speaker. Producing it in a non-content turn means that there is no opportunity to explain it, which again increases the cost. As discussed in section 2.1.1.4, the only occurrence of a participant making a non-content disagreement turn to the moderator was where they were using humour and sarcasm to try and downplay their dispreferred response

3. Response to Moderator: Content Disagreement Turn

This turn type would also carry a high potential cost to the speaker. Not only are they giving a dispreferred response but it also to a more powerful speaker. Being a content turn does allow for the opportunity to justify the dispreferred response and for this reason it would carry less inherent strength than the non-content turn above. Perhaps because of the associated cost of such an evaluative turn, there were no examples of this within my research

4. Response to Participant: Non-content Disagreement Turn

As with the non-content disagreement turn to the moderator, this turn type would carry a heavy cost to the speaker. However in comparison it is less indicative of a strong evaluation due to the lack of power difference between speaker and receiver. Again there were no incidences of this turn type within my research data.

5. Response to Participant: Content Disagreement Turn

The strong structural forces that impact on all conversations means that providing a dispreferred response constitutes a face threatening act. The additional cost to the speaker of making this turn type marks it out as being indicative of a stronger evaluation.

6. Response to Moderator Question: Content Turn

I argue that evaluations contained within this turn type are neither inherently strong nor inherently weak. Though research effects might affect respondents' attitudes towards the moderator and the stimulus, the notion of preference should be less relevant here meaning there is less encoded evaluative force. The content nature of the turn means that it is not indicative of a weaker evaluation and the fact that it is responding to a more powerful speaker means that it is not indicative of a stronger evaluations. Therefore this turn type is viewed as the midpoint between those structures indicating high and low inherent evaluative force.

7. Response to Participant: Content Agreement Turn

Making this kind of turn carries little risk to the speaker; they have been structurally invited to comment and they are agreeing with what has previously been said so there is no element that constitutes a face threatening act. The content element means that it is also complying with Grice's maxim of quantity.

8. Response to Participant: Non-content Agreement Turn

This turn type is indicative of minimal force from the speaker and as suggested in section 2.1.1.4, could even be considered as lacking in evaluation depending on phonological aspects of its production.

9. Response to Moderator: Content Agreement Turn

As discussed in section 2.1.1.3.1, the only occasions where it is likely that there will be an explicit preferred response to a moderator assessment is where the moderator is summarising or clarifying previous points made by the respondents. Producing this turn type has little inherent cost as it has been invited by a more powerful speaker and simply agrees with it.

10. Response to Moderator: Non-content Agreement Turn

I argue that the cost of making this turn is minimal for the participant as they are simply agreeing with a speaker who has a more powerful role in the interaction. The non-content aspect has two different roles that may further minimise the cost of the turn and the implied strength of the evaluation. Firstly, it shows that they are not attempting to alter the assessment in any way with the implication being that they completely agree. Secondly it may be that they are not particularly interested in this assessment at all and are using their weak agreement to take control of the floor.

This scale maps onto the Turn Taking layer used in the CorpusTool analysis. However, while the turn taking categories are ranked according to the perceived 'cost' of making the evaluation, this scale falls outside other markers of strength from the Appraisal layer such as FORCE or FOCUS markers and these are dealt with separately in the Appraisal Layer.

5.4.3 The Participant Layer

As noted in the introduction to this section, this layer simply shows which of the participants in the communication was responsible for each utterance. Therefore each turn was tagged as belonging to either the Moderator, Speaker A or Speaker B.

Having discussed the reasons for the way in which the data was tagged, the next section will detail the inter-rater reliability testing that was carried out.

5.5 Inter-rater Reliability

One possible criticism of a methodology that uses an approach such as Appraisal Theory is that it is open to significant variance between different practitioners. This means that two researchers working with the same data might come to different conclusions based on their tagging and analysis of the data. For a methodology that uses categories with fuzzy boundaries this is, to a certain extent, an unavoidable problem. However as Page (2003: 216) states:

Like all analysis of appraisal, but perhaps even more so, the classification of these evocations involves a degree of subjectivity. As such, these instances present occasions where there may be the possibility for multiple interpretations.

That there may be fuzzy boundaries or alternative readings does not necessarily render the methodology obsolete. Macken-Horarik (2003: 316) suggests that any type of coding is potentially problematic arguing that any 'coding of an item (word or wording) is itself a fragile process'. There are two different approaches to minimising the effect of this problem. Firstly it is essential that the researcher is consistent within their own research project. Where there are over-lapping boundaries and subjective decisions to be made about certain data items, it is important to ensure that the analysis has an internal reliability. As long as the context does not alter the meaning, each instance of the same

utterance or expression should be tagged in the same way. To help ensure that this consistency was maintained, evaluations were compared to a database of evaluations that was compiled throughout the tagging process. This database contained each of the different evaluative units that occurred and detailed which of the categories of the Appraisal Theory framework they had been assigned to. The CorpusTool software allows for searches by specific words and this means that it is possible to check that they have each been tagged in the same way. There will of course be cases where the same lexeme has been tagged in different ways depending on its part of speech or local context as explored above in section 5.4.1.1.5. Macken-Horarik (2003: 316) shows the importance of this flexibility suggesting that without it ‘coding requires that we settle on a valeur for any item of APPRAISAL in a text’.

The second way in which this potential problem can be minimised is through testing for inter-rater reliability. This is a process whereby a number of analysts are given the same data and analyse it using the same pre-agreed set of categories. Silverman (2006: 288) argues that ‘reliability can be improved by comparing the analysis of the same data by several researchers.’ Where there are differences, these can then be discussed and any underlying causes can be resolved. However I would argue that it would be expected that there might be differences between the codings of different researchers due to their different perspectives and approaches to the data. Macken-Horarik (2003) emphasises the importance of considering the analysts’ perspective as this may have an impact on the way that they code the data. Hood (2004: 113) argues that ‘while the analyses in this study are undertaken on a theoretically principled basis, a degree of subjectivity is none the less intrinsic to a study of attitudinal meaning’. That different readers may read the texts from different subject positions is also a view taken by Kress (1985) and Martin (1995). Coding may also be changed according to whether the analyst tags as they go along or reads ahead further to gain insights into where the evaluation is heading and how it fits into the broader discourse. These differing approaches are labelled as ‘dynamic’ or ‘synoptic’ (Macken-Horarik 2003).

5.5.1 Testing Reliability

To test the reliability of my data tagging, two sections of transcripts were tagged by two other researchers. Although not practicing Appraisal Theorists, they each have post-graduate qualifications in Applied Linguistics and are familiar with the processes involved in transcribing, coding and analysing data, particularly with regards to a theory based on Systemic Functional Grammar. Each of the researchers was sent additional information to provide the context for the research and to ensure that they had sufficient knowledge of Appraisal Theory to confidently code the sections of transcripts that were provided. Although it might be argued that the context should not be important if there are clearly defined categories and boundaries, I have argued elsewhere that both the local and general context can be important in helping to ascertain the correct tagging of certain evaluations. In addition to providing the context for this research, the researchers were also provided with the relevant sections on Appraisal Theory from this thesis. Two sections of transcript were sent from different stimuli and different interviews. The sections of transcript that were tested for inter-rater reliability totaled approximately thirty minutes of data and included over one hundred different tags. The full information pack and guidelines that were provided can be seen in Appendix IV.

5.5.2 Comparing the Tagging

Inter-rater reliability on the two examples transcripts was assessed using the Kappa statistic (Cohen, 1960; Siegel & Castellan, 1988; Carletta, 1996). Shriberg (1998) describes this as the 'ratio of the proportion of times that raters agree (corrected for chance agreement) to the maximum proportion of times that the rates could agree (corrected for chance agreement)'.

Kappa computed for the rating of the two sets of annotations was 0.71. This is suggested as showing substantial agreement using Landis and Koch's (1977) scale where .21-.40 is fair, .41-.60 is moderate, .61-.80 is substantial, and .81-1 is almost perfect agreement. There are two factors that had a particular impact on this figure. Firstly, the majority of

the differences occurred where the co-raters did not apply any tags, rather than applying the wrong ones. Many of these instances were caused by inherent evaluations in content agreement turns being tagged by myself but not by the co-rater. For example:

166. Speaker A: *That's just too bulky.*
167. Speaker B: *Yea it is definitely yea.*

Here both initial evaluation and the response were tagged as COMPOSITION: BALANCE: NEGATIVE but the co-raters only tagged the initial evaluation and not the inherent evaluation in the response. The other main cause of difference was with regards to the ENGAGEMENT branch of the framework, particularly those following under the sub-branch of PROCLAIM. The potential for different interpretations for evaluations of this kind is discussed in more detail in section 5.4.1.1.4.

This chapter set out to outline and justify the data collection methods that were used in this research in addition to discussing the transcribing and tagging of the data once it had been collected. Having done this, the following chapter will display the data and introduce a preliminary discussion of certain points of interest.

6. Results

Having detailed the research methodologies that were used to collect the data in the previous chapter, this chapter presents the data that is the basis for the discussion in the following chapter. As well as describing the data, this chapter also previews certain points of interest. These points of interest are then carried forward into the main discussion chapter. As noted above, although there were five stimuli used in each paired depth interview, I will only be using the data from two of these stimuli. The two that have been chosen are the most and least liked packs from the explicit ranking exercises that were carried out at the end of each interview. This means that there should be a full spread of both positive and negative language pertaining to evaluation. These two stimuli were named Mouse (Figure 25) and Egypt (Figure 26), with Egypt being the least liked and Mouse the most liked from the explicit ranking exercise. Each of these will be explored separately and then a comparison will be made between the two sets of data. Mouse was so called as it was felt to resemble a computer mouse and Egypt was given this name as in the initial designs the respondent had talked about it having an Egyptian theme.



Figure 25 Stimulus Mouse



Figure 26 Stimulus Egypt

I will start off by showing the way in which positive and negative evaluations were distributed across the different parts of the Appraisal Theory Framework. This will help to show general trends in terms of the way in which the participants evaluated the different stimuli. It will also help to highlight any differences between the explicit rankings that were given to each stimulus and the less explicit evaluations that were made

in the paired depth interviews. While contrasting positive and negative evaluations might prove fruitful in terms of highlighting key points for discussion, it does presuppose that any evaluation is either positive or negative. This leaves little room for the notion of a neutral evaluation and this is a discussion that is introduced in this section before being explored in more detail in the main Discussion Chapter that follows.

In both this chapter and chapter 7, percentages are compared for statistical significance using a z-test with a confidence level of 95%. Therefore any calculation with a p-level of less than 0.05 will be considered to be statistically meaningful and those calculations with a p-value of greater than 0.05 will not be considered statistically meaningful.

In section 6.1, I present the data according to its polarity and place within the Appraisal Theory Framework. The evaluations are categorised with regards to each of the different sub-sets of the ATTITUDE branch of framework. This includes evaluations relating to the COMPOSITION, REACTION, VALUATION, AFFECT and JUDGEMENT of the stimuli. Section 6.2 then details the way in which the evaluations are weakened or strengthened through the use of specific, lexical weakening or strengthening strategies. These are categorised with regards to each of the different sub-sets of the GRADUATION and ENGAGEMENT branches of the framework. Based on data tagged and analysed under the ENGAGEMENT part of the Appraisal Theory Framework, this section also introduces the suggestion of potential developments and extensions of the framework. These extensions are with regards to the relevance of the subject matter to the speaker making the evaluation, and the notion of neutral evaluations. These ideas are explored in further detail within the Discussion Chapter that follows. Section 6.3 examines the inherent weakening and strengthening that is attached to an evaluation as a result of its place in the turn-taking structure. This draws on the framework discussed in detail in section 5.4.2.

6.1 *Distribution of Positive and Negative Evaluations across Appraisal Theory Framework*

This section will look at the distribution of positive and negative evaluations of both Egypt and Mouse. Presenting the data in this way will act as a precursor to answering the final research sub – questions:

- 1.c) Is there a link between the overall ranking of stimulus and the distribution of positive and negative evaluations?
- 2.a) Are different categories of the Appraisal Theory Framework more likely to attract positive or negative evaluations and does this have consequences with regards to the participants' overall evaluation of the pack?

A table will be used to show the total number of evaluations that were made within each area of the Appraisal Theory Framework. This will then be exemplified with excerpts from the transcripts of the paired depth interviews. Some of the evaluations were tagged for more than one part of the framework, which means that the total number of evaluations might be more than the sum of the different parts of the framework. This occurred when an evaluation was tagged for either positive or negative AFFECT and then also for the branch of the framework that was relevant to the cause of the AFFECT, as in the following examples

168. Speaker A: *I don't like the erm button on the top.*

This utterance was tagged for both NEGATIVE AFFECT and NEGATIVE COMPOSITION: BALANCE as the unhappiness with the stimulus was caused by the placement of the button being asymmetric and discordant. The emotional response of liking or not liking the stimulus is captured within the AFFECT branch of the Appraisal Theory Framework. However, if this utterance was only tagged in this way it would seem to miss some of the focus of the evaluation. Therefore the target of the AFFECT is also tagged to try and

account for the full evaluative content of the utterance. As discussed in section 5.4.1.1.1, this process of double tagging is sometimes necessary to capture the different levels of evaluation which can be contained within a single utterance. This double tagging occurred across each of the different evaluation types, regardless of polarity and is therefore relevant for all of the tables in sections 6.1.1 to 6.1.4.

This section will describe the following four evaluation types:

- Positive evaluations about Egypt
- Positive evaluations about Mouse
- Negative evaluations about Egypt
- Negative evaluations about Mouse

There will then be a brief comparison and discussion of the two data sets highlighting some of the main issues that will be discussed in more detail in the Discussion Chapter. In total there 509 evaluation about Egypt and 765 about Mouse giving a total number of evaluations of 1274.

6.1.1 Positive Evaluations for Egypt

In total there were 132 positive evaluations for the Egypt pack across the different categories of the ATTITUDE branch of the Appraisal Theory Framework. This means that 26% of all evaluations about Egypt were positive. Table 10 shows their distribution.

	Egypt	%
Composition	26	19.9
Reaction	50	37.8
Valuation	43	32.6
Affect	9	7.4
Judgement	4	2.9
Total	132	

Table 8 Distribution of Positive Evaluations about Egypt Across the Appraisal Theory Framework

This table shows that 71% of the positive evaluations were made with regards to the initial reaction that the pack caused and the overall valuation that followed. There were comparatively few evaluations with regard to the pack's normality or the emotional response that it triggered.

Examples of positive COMPOSITION are:

- 169. *It's a nice big size.*
- 170. *Yea yea good to have grips yea.*
- 171. *I quite like the er the groove thing.*

Examples of positive REACTION are:

- 172. *It would probably stand out.*
- 173. *Aesthetically it looks quite nice pleasing.*
- 174. *Yea I think I would notice it.*

Examples of positive VALUATION are:

- 175. *It's new it's a new shape.*
- 176. *It does look a bit different.*
- 177. *I would have said more modern.*

Examples of positive AFFECT are:

- 178. *Yea I like that.*
- 179. *I quite like to see that.*

Examples of positive JUDGEMENT are:

- 180. *It's not that different really.*

181. *So it's kind of the same shape.*

6.1.2 Positive Evaluations for Mouse

The participants made 474 positive evaluations about the stimulus Mouse across the different categories of the ATTITUDE branch of the Appraisal Theory Framework. Out of a total of 765 evaluations this means that 62% were positive. In comparison only 26% of evaluations about Egypt were positive which is statistically significant to a value below 0.001. The breakdown is shown in Table 11.

	Mouse	%
Composition	68	14.3
Reaction	152	32.1
Valuation	192	40.5
Affect	54	11.4
Judgement	8	1.7
Total	474	

Table 9 Distribution of Positive Evaluations about Mouse Across the Appraisal Theory Framework

This table shows that the distribution of positive evaluations was similar for both stimuli. Whereas with Egypt, 71% of the positive evaluations were in relation to the initial reaction that it created and the overall valuation that followed, for Mouse this figure was 73% which means that they were not significantly different. This would be of particular interest to brand owners or product designers seeking to create a product with high impact or 'visibility', but who were not as concerned with the product's ease of use. Although usability needs are seen as more important than pleasure needs (see figure 2), for low involvement products such as deodorant it is possible that a brand could choose high shelf presence and a high efficacy of the product over pure ergonomics and ease of use.

Examples of positive COMPOSITION are:

182. *It feels quite easy to hold.*

183. *It's a better size yea.*
184. *It just fits perfect.*

Examples of positive REACTION are:

185. *I just think it's a very feminine shape.*
186. *Oh it would stand out.*
187. *I like the design I think it looks nice.*

Examples of positive VALUATION are:

188. *If it was on the shelf I would try it.*
189. *It's intriguing I think.*
190. *I would I would certainly want to try that one.*

Examples of positive AFFECT are:

191. *I do like that.*
192. *I quite like the different button on it.*

Examples of positive JUDGEMENT are:

193. *It still looks like a deodorant.*
194. *I haven't seen anything like that before.*

The fact that there were over 3.5 times more positive evaluations for Mouse than for Egypt is something that will be discussed further in section 7.1. Assessing the use of weakening and strengthening techniques within these evaluations will also help explore any link between the rankings from the explicit ranking exercise and the less explicit evaluations from the paired depth interviews. These techniques will be explored after the distribution of the negative evaluations has also been presented and discussed below.

6.1.3 Negative Evaluations for Egypt

In total there were 377 negative evaluations made about Egypt across the different categories of the ATTITUDE branch of the Appraisal Theory Framework. This means that 74% of evaluations about Egypt were negative.

	Egypt	%
Composition	174	46.1
Reaction	110	29.2
Valuation	47	12.5
Affect	31	8.2
Judgement	15	4
Total	377	

Table 10 Distribution of Negative Evaluations about Egypt Across the Appraisal Theory Framework

Table 12 shows that 46% of the negative evaluations made about Egypt were with regards to the design and ease of use of the pack. A further 29% of the evaluations were related to the initial impact and the quality of the pack. Only 4% of the negative evaluations were with regards to JUDGEMENT: NORMALITY. This might suggest that although the pack was fairly consistent with the expectations of the participants in terms of what a deodorant pack should look like, there was something else about it that caused the negative evaluations. The high proportion of negative evaluations with regard to the COMPOSITION of the pack suggests that it was the right kind of shape but in the wrong dimensions. Some support for this interpretation can be seen with utterances such as:

195. *I like the idea if it were smaller.*
196. *A third of the size.*
197. *I think the dimensions are far too big.*

In each of the examples the criticism is aimed specifically at the proportions rather than the design or ease of use or normality. There are subsequent negative evaluations that relate to ease of use and normality but again these are caused by the dimensions rather than by complicated or distinctive designs.

198. *It's a little bit too big in the hand.*
199. *Don't tend to get them as that big normally.*

Examples of negative COMPOSITION are:

200. *It's too big.*
201. *It's a bit big and bulky.*
202. *It just it doesn't it doesn't sit nicely in the hand.*

Examples of negative REACTION are:

203. *I'd not notice it at all.*
204. *Just a bit of an eyesore for me.*
205. *And it's ugly.*

Examples of negative VALUATION are:

206. *I just see it as gimmicky.*
207. *There's nothing to it.*
208. *It just looks a bit unfinished to me.*

Examples of negative AFFECT are:

209. *No I don't like that at all.*
210. *Yea probably embarrassed carrying it.*
211. *I'm not keen on it.*

Examples of negative JUDGEMENT are:

212. *It's a bit bigger than usual ones isn't it.*

213. *It is something out of the ordinary isn't it it's not something I've seen before.*
214. *We're used to more sort of thinner canisters.*

6.1.4 Negative Evaluations for Mouse

Although Mouse was the most favoured stimulus from the ranking exercises, it still received many negative evaluations across the different categories of the ATTITUDE branch of the Appraisal Theory Framework.

	Mouse	%
Composition	124	42.6
Reaction	56	19.2
Valuation	68	23.4
Affect	33	11.3
Judgement	10	3.4
Total	291	

Table 11 Distribution of Negative Evaluations about Mouse Across the Appraisal Theory Framework

In total there were 291 (38%) negative evaluations for Mouse, in comparison to 377 (74.1%) for Egypt which is statistically significant to a value below 0.001. 42.6% of the negative evaluations about Mouse were with regards to the design and ease of use of the pack which is not significantly different from the 46% of negative evaluations about Egypt made up of the same category. The proportion of negative evaluations about Mouse relating to the pack's normality (JUDGEMENT), 3.4%, or the emotional response that it triggered (AFFECT), 11.3%, were also significantly similar to that of Egypt at 4% and 8.2% respectively.

Examples of negative COMPOSITION are:

215. *Bit big though isn't it.*
216. *It's not obvious is it.*
217. *It sort of doesn't stand up or anything.*

Examples of negative REACTION are:

- 218. *I think first impression is I don't like it but who knows really I suppose.*
- 219. *I don't know it just doesn't appeal to me really.*
- 220. *I can't see it being pretty that shape.*

Examples of negative VALUATION are:

- 221. *I don't think I'd go for something quite that big.*
- 222. *Yea posery designer type.*
- 223. *It's a little bit gimmicky.*

Examples of negative AFFECT are:

- 224. *I don't like it.*
- 225. *No I wouldn't say I was particularly keen.*
- 226. *I always worry that they're going to spray everywhere.*

Examples of negative JUDGEMENT are:

- 227. *It's a bizarre shape for a deodorant isn't it.*
- 228. *It's just really odd.*
- 229. *I think it's more a change of what you're used to.*

Analysing the distribution of positive and negative evaluations can be used to give an indication as to the participants' attitudes and to spot any trends or patterns. However this simplistic analysis can be strengthened by also taking into account the use of any weakening and strengthening features displayed by the speakers. The use of such strategies is displayed below and discussed in more detail in section 7.3.

6.2 Weakening and Strengthening of Evaluations in Relation to their Polarity

Another area of interest is in the patterns of usage with regards to the polarity of an evaluation and the frequency and type of strengthening or weakening markers that were used. Examining the data in this way will be the first stage in answering the remaining research sub – questions:

- 1.a) Is there a link between the distribution of strengthening or weakening markers and the explicit ratings given in the ranking exercise? (I.e. were participants more likely to strengthen negative evaluations of the least liked pack or strengthen positive evaluations of the most liked pack?)

- 1.b) Is there a link between the polarity of an evaluation and the distribution of strengthening or weakening markers? (I.e. are negative evaluations more likely to be weakened because of, say, politeness reasons?)

Firstly this section will look at the relative weakening and strengthening of positive evaluations across both stimuli before doing the same for negative evaluations. This section will be comprised of the following eight areas:

- Weakening of positive evaluations about Egypt
- Strengthening of positive evaluations about Egypt
- Weakening of positive evaluations about Mouse
- Strengthening of positive evaluations about Mouse
- Weakening of negative evaluations about Egypt
- Strengthening of negative evaluations about Egypt
- Weakening of negative evaluations about Mouse
- Strengthening of negative evaluations about Mouse

6.2.1 Weakening of Positive Evaluations about Egypt

Of the 132 positive evaluations that were made about Egypt, 58% (77) used one or more of the possible weakening strategies to downplay the strength of their positive evaluation. In Table 14, the rows show the different weakening strategies that were used and the columns show the different parts of the Appraisal Theory Framework.

	Composition Balance (N=15)		Composition Complexity (N=12)		Reaction Impact (N=14)		Reaction Quality (N=36)		Valuation (N=46)		Judgement (N=6)		Affect (N=11)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Force (Lower)	3	20	1	8	1	7	12	33	6	13	1	20	4	36
Focus (Soften)	0	0	0	0	1	7	8	22	8	18	1	20	0	0
Engagement (Decrease)	2	13	2	16	9	64	19	55	15	31	2	40	1	9
No Weakening Strategy	11	73	3	75	4	28	9	25	23	47	2	40	7	63

Table 12 Distribution of different weakening strategies for positive evaluations about Egypt

Table 10 in section 6.1.1 showed that 71% of the positive evaluations for Egypt came under the REACTION and VALUATION branches of the framework. However, this table shows that these categories were also amongst the most likely to have their positive evaluations weakened through the use of GRADUATION and ENGAGEMENT. In contrast, although there were approximately half as many positive evaluations regarding the COMPOSITION of the packs, these were far less likely to have been weakened. All three options, from the GRADUATION and ENGAGEMENT branches of the framework, for mitigating or hedging an evaluation were used.

Fifty evaluations were weakened by ENGAGEMENT: DECREASE:

230. *Yea I think they'd be good yea.*

231. *So I reckon that be a seller.*

232. *Yea I probably would look at it.*

Twenty-eight were weakened by FORCE: LOWER;

- 233. *Looks wise I think it is **quite** a nice design.*
- 234. *It does look **a bit** different.*
- 235. *Nice **little** touch.*

Eighteen were weakened by FOCUS: SOFTEN:

- 236. *So it's **kind of** the same shape.*
- 237. *It **sort of** breaks away from the norm.*
- 238. *I would have said **more** modern.*

In total only 45% (59) of positive evaluations were not weakened in some way. This figure includes positive evaluations which used the same strategies of FORCE FOCUS and ENGAGEMENT to increase the strength of their evaluation.

6.2.2 Strengthening of Positive Evaluations about Egypt

58% (77) of the positive evaluations for the Egypt pack were mitigated or hedged in some way but only 11% (14) of the positive evaluations were strengthened through GRADUATION or ENGAGEMENT which is significantly different.

Of these fourteen occurrences, twelve used an increase of ENGAGEMENT, one used a heightening of intensity of FORCE and one used both. There were no examples of positive evaluations that sharpened the FOCUS of the evaluation.

	Composition Balance (N=15)		Composition Complexity (N=12)		Reaction Impact (N=14)		Reaction Quality (N=35)		Valuation (N=44)		Judgement (N=5)		Affect (N=11)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Force (Heighten)	0	0	0	0	0	0	1	3	1	2	0	0	0	0
Focus (Sharpen)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Engagement (Increase)	1	7	2	17	0	0	6	17	6	14	1	20	1	9
No Strengthening Strategy	14	93	10	83	14	100	28	80	37	84	4	80	10	91

Table 13 Distribution of different strengthening strategies for positive evaluations about Egypt

Table 15 shows that 88.6% of Positive Evaluations about Egypt did not make use of any of the available strengthening strategies under the Appraisal Theory Framework. The rows show the different strengthening strategies that were used and the columns show the different parts of the Appraisal Theory Framework. The distribution of strengthening strategies across the different parts of the Appraisal Theory Framework was fairly equal, ranging from no strengthening strategies with regards to the IMPACT of the stimulus to 20% of the positive evaluations being strengthened for both the QUALITY and the JUDGEMENT of the stimulus. Where they were strengthened, it was usually through ENGAGEMENT: INCREASE with no examples of FOCUS: SHARPEN and only two examples of FORCE: HEIGHTEN.

Examples of ENGAGEMENT: INCREASE are:

- 239. *Certainly wouldn't think it were cheap.*
- 240. *Yea definitely again and again yea*

Examples of FORCE: HEIGHTEN are:

- 241. *I think the ridges the slight ridges on the side are a good idea.*
- 242. *I would probably say that it would be more expensive.*

Having looked at the way in which positive evaluations were weakened or strengthened for Egypt, the following two sections will compare this to the way in which the participants weakened or strengthened these positive evaluations for Mouse. In total there were 474 positive evaluations made about the stimulus Mouse and these were spread out across all parts of the Appraisal Theory Framework.

6.2.3 Weakening of Positive Evaluations about Mouse

Of the 474 positive evaluations about Mouse, 210 (44.3%) were weakened using one or more of the available strategies from GRADUATION or ENGAGEMENT. This compares with 55% of the positive evaluations that were weakened for Egypt and is statistically significant.

	Composition Balance (N=68)		Composition Complexity (N=10)		Reaction Impact (N=20)		Reaction Quality (N=132)		Valuation (N=194)		Judgement (N=8)		Affect (N=54)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Force (Lower)	6	9	2	20	0	0	20	15	36	19	0	0	22	36
Focus (Soften)	2	3	2	20	0	0	38	29	14	7	0	0	0	0
Engagement (Decrease)	4	6	0	0	0	0	52	39	70	36	4	40	6	9
No Weakening Strategy	56	88	6	60	20	100	60	45	104	54	4	50	30	56

Table 14 Distribution of different weakening strategies for positive evaluations about Mouse

Table 16 shows that although there were more positive evaluations containing REACTION and VALUATION, these categories were more likely to have been weakened than other categories. There were four positive evaluations concerning REACTION and VALUATION (71%) for each one relating to COMPOSITION (17%) but whereas 47% of these were weakened for REACTION and VALUATION, only 20% were weakened for COMPOSITION which is statistically significant.

55% of evaluations containing JUDGEMENT or AFFECT were weakened but while evaluations containing AFFECT were mostly weakened through FORCE: LOWER,

evaluations containing JUDGEMENT were always weakened through ENGAGEMENT: DECREASE.

Examples of FORCE: LOWER are:

- 243. *I quite like that one.*
- 244. *A bit funky and a bit different.*

Examples of FOCUS: SOFTEN are

- 245. *It's sort of modern.*
- 246. *It's like a young design a trendy design.*

Examples of ENGAGEMENT: DECREASE are:

- 247. *I'd probably buy it.*
- 248. *It reminds me of a candle.*
- 249. *I think that'd be quite handy for the bathroom.*

The strengthening strategies were limited to REACTION: IMPACT, REACTION: QUALITY and VALUATION. In contrast the weakening strategies were used in each part of the framework apart from REACTION: IMPACT.

Examples of COMPOSITION: BALANCE are:

- 250. *It feels quite easy to hold.*
- 251. *It looks like you can get a good grip on it.*

Examples of COMPOSITION: COMPLEXITY are:

- 252. *That would be quite functional like that.*

253. *More sort of practical and like you say functional.*

Examples of REACTION: QUALITY are:

254. *It's quite pretty.*

255. *I think more the shape really it's more decorative for dressing tables.*

256. *It looks more of a like Armani type or a Hugo Boss.*

Examples of VALUATION are:

257. *Yea I just think it'd be a bit of a talking point I think.*

258. *I think it's interesting.*

259. *I'd still probably use that.*

Examples of JUDGEMENT are:

260. *I don't think I've ever seen one that's laid flat.*

261. *So it's probably different.*

Examples of AFFECT are:

262. *I quite like that one.*

263. *I'd be quite happy to display that cos again you could imagine that being a bit of a feature.*

Having looked at the way in which positive evaluations about Mouse were weakened, the following section will look at the different ways in which they were strengthened.

6.2.4 Strengthening of Positive Evaluations about Mouse

In total, of the 474 positive evaluations about Mouse, 76 (16%) were explicitly strengthened through the use of increased ENGAGEMENT or heightened FORCE. This does not appear to be statistically different from the 11% of positive evaluations that were strengthened for Egypt and this is something that is revisited in section 7.2.2. There were no examples of a positive evaluation that made use of the FOCUS: SHARPEN part of the Appraisal Framework. Examples of FORCE and ENGAGEMENT being used to strengthen evaluations are shown below.

	Composition Balance (N=68)		Composition Complexity (N=10)		Reaction Impact (N=20)		Reaction Quality (N=132)		Valuation (N=194)		Judgement (N=8)		Affect (N=54)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Force (Heighten)	0	0	0	0	2	10	18	7	14	2	0	0	0	0
Focus (Sharpen)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Engagement (Increase)	0	0	0	0	2	10	20	17	22	14	0	0	0	0
No Strengthening Strategy	68	100	10	100	16	80	102	77	158	81	8	100	54	100

Table 15 Distribution of different strengthening strategies for positive evaluations about Mouse

Table 17 shows that although 16% of the positive evaluations were strengthened, the distribution was not spread out amongst the different parts of the Appraisal Theory Framework. The rows show the different strengthening strategies that were used and the columns show the different parts of the Appraisal Theory Framework. Only evaluations relating to the QUALITY and overall VALUATION of the stimulus were strengthened and only FORCE: HEIGHTENING and ENGAGEMENT: INCREASE were used. There were no examples of FOCUS: SHARPEN.

Examples of FORCE: HEIGHTEN are:

264. *It's so so novel.*

265. *It's a very feminine shape.*

266. *yea I really like that.*

Examples of ENGAGEMENT: INCREASE are:

267. *I'd certainly try that one.*

268. *I'd look at it definitely.*

269. *I'd say a pebble off a beach type thing.*

Although there were positive evaluations across each of the different parts of the Appraisal Framework, the distribution of those that were strengthened was far from equal. While there were no positive evaluations regarding COMPOSITION: BALANCE, COMPOSITION: COMPLEXITY, JUDGEMENT or AFFECT that had been strengthened, 21% of the positive evaluations relating to REACTION: IMPACT, REACTION: QUALITY and VALUATION were strengthened in some way. This distribution might support the idea, suggested in section 6.1.2, that the participants initial reaction and overall valuation of the pack was strong enough that their negative evaluations of the actual dimensions and ease of use of the pack were less important to them.

Having looked at the way in which positive evaluations about Mouse and Egypt were weakened or strengthened, I will now do the same for negative evaluations.

6.2.5 Weakening of Negative Evaluations about Egypt

Of the 377 negative evaluations about Egypt, 47% (177) made use of one or more of the possible weakening strategies to downplay the strength of the negative evaluation.

	Composition Balance (N=147)		Composition Complexity (N=30)		Reaction Impact (N=27)		Reaction Quality (N=84)		Valuation (N=50)		Judgement (N=15)		Affect (N=32)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Force (Lower)	43	29	11	37	8	30	13	16	10	20	2	13	3	9
Focus (Soften)	6	4	1	3	0	0	22	26	6	12	2	13	0	0
Engagement (Decrease)	27	19	10	33	8	30	29	35	21	42	2	13	5	16
No Weakening Strategy	87	59	11	37	14	52	38	45	21	42	9	60	25	78

Table 16 distribution of different weakening strategies for negative evaluations about Egypt

Table 18 shows that the distribution of the weakening strategies for negative evaluations was fairly equal across each of the different categories of the Appraisal Theory Framework. The rows show the different weakening strategies that were used and the columns show the different categories. Negative evaluations containing AFFECT were the least likely to have been weakened and evaluations containing COMPOSITION: COMPLEXITY were most likely to have been weakened. The three main options for mitigating or hedging an evaluation were all used.

Examples of ENGAGEMENT: DECREASE are:

- 270. *It's the diameter it's too wide and **probably** a bit too tall as well.*
- 271. *I don't **think** I'd use it personally.*
- 272. *It just **looks** too big for me yea.*

Examples of FORCE: LOWER are:

- 273. *It's **quite** big.*
- 274. *It's a **little** bit too wide.*
- 275. *I'm **slightly** concerned that the spray mechanism's exposed.*

Examples of FOCUS: SOFTEN are:

276. *Mainly size thickness height.*
 277. *We're used to more sort of thinner canisters.*
 278. *It's like a value pack.*

54% of the negative evaluations about Egypt made use of one or more of the different approaches that are available to a speaker to mitigate or hedge the strength of their opinion or evaluation. This is not statistically different from the 44% of negative evaluations about Mouse that were also weakened. This is discussed in greater detail in section 7.2.3 below. The next section will look at the way in which some negative evaluations were strengthened by the speaker. If the pattern of finding negative evaluations easier to make than positive ones is consistent, then the speakers should be more likely to strengthen their negative evaluations than their positive evaluations. This will be discussed below.

6.2.6 Strengthening of Negative evaluations about Egypt

While a close analysis of the positive evaluations about Egypt suggested that there were few examples that had been strengthened in a clear and unproblematic way, the same is not true for negative evaluations. Of the 377 negative evaluations about Egypt, 67 (18%) made use of one or more of the possible strengthening strategies to increase the strength of the negative evaluation.

	Composition Balance (N=147)		Composition Complexity (N=30)		Reaction Impact (N=27)		Reaction Quality (N=84)		Valuation (N=50)		Judgement (N=15)		Affect (N=32)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Force (Heighten)	15	10	4	13	1	4	10	12	3	6	0	0	3	10
Focus (Sharpen)	2	1	0	0	0	0	0	0	2	4	0	0	0	0
Engagement (Increase)	14	9	0	0	1	4	9	11	3	6	0	0	2	6.2
No Strengthening Strategy	126	86	26	87	25	93	66	79	42	84	15	100	27	84

Table 17 Distribution of different strengthening strategies for negative evaluations about Egypt

Table 19 shows that the distribution of strengthening strategies was very similar across the different categories of the Appraisal Theory Framework. The rows show the different strengthening strategies that were used and the columns show the different categories. Negative evaluations containing COMPOSITION, REACTION, VALUATION or AFFECT were strengthened 13%, 18%, 16% and 14% respectively. Only negative evaluations containing JUDGEMENT deviated from this pattern as none of these were strengthened at all. In total 18% of negative evaluations were strengthened in some way by the speaker and this in comparison to 11% for positive evaluations. While the difference highlighted by this surface level comparison is not statistically significant it does not take into account the distribution of additional strategies for strengthening or weakening evaluations and this is discussed in more detail in section 7.2. Both GRADUATION and ENGAGEMENT were used to increase the strength of the negative evaluations:

Examples of INTENSITY: HEIGHTEN are:

- 279. *Very old fashioned.*
- 280. *A really weird shape.*
- 281. *It's far too big.*

Examples of FOCUS: SHARPEN are:

- 282. *It's like a real brick.*
- 283. *It's just the actual dimensions I think are too bulky to use.*
- 284. *The actual holding of it cos that is awkward.*

Examples of ENGAGEMENT: INCREASE are:

- 285. *Yea I'd say it were too big.*
- 286. *Well I personally don't like it.*
- 287. *Definitely wouldn't buy it.*

Having shown how negative evaluations about Egypt were weakened and strengthened, the following sections will do the same for Mouse.

6.2.7 Weakening of Negative Evaluations about Mouse

	Composition Balance (N=104)		Composition Complexity (N=20)		Reaction Impact (N=4)		Reaction Quality (N=52)		Valuation (N=68)		Judgement (N=10)		Affect (N=32)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Force (Lower)	34	33	6	30	2	50	4	8	12	18	4	20	0	0
Focus (Soften)	6	6	2	10	2	50	28	54	10	15	0	0	0	0
Engagement (Decrease)	26	25	10	50	4	100	24	47	34	50	0	0	0	0
No Weakening Strategy	56	54	8	40	0	0	16	31	24	35	8	80	100	100

Table 18 Distribution of different weakening strategies for negative evaluations about Mouse

Of the 291 negative evaluations about Mouse, 51% (148) made use of one or more of the possible weakening strategies to downplay the strength of the negative evaluation. Table 20 shows that the distribution of weakening strategies for negative evaluations was different across each of the categories of the Appraisal Theory Framework. The rows show the different weakening strategies that were used and the columns show the different categories. For REACTION: IMPACT all of the evaluations were weakened in some way and for negative evaluations containing AFFECT, the opposite was true, with none of the evaluations being weakened. Each of the options available within the GRADUATION and ENGAGEMENT branches of the Appraisal Theory Framework was used.

Examples of FORCE: LOWER are:

- 288. *Bit big though isn't it.*
- 289. *It's still quite wide.*
- 290. *It feels a little unnatural.*

Examples of FOCUS: SOFTEN are:

- 291. *That looks like an air freshener.*

292. *Sort of this end here is quite fat.*

293. *More of a poser type.*

Examples of ENGAGEMENT: DECREASE are:

294. *I don't think it'd take off.*

295. *I probably wouldn't get that.*

296. *You'd have to get used to it maybe.*

As the most liked pack, it might be expected that there would be fewer negative evaluations or that a higher percentage of them would also be weakened. However there were a similar number of negative evaluations. The occurrence of negative evaluations that were both weakened and strengthened is described below in section 6.2.8.

6.2.8 Strengthening of Negative Evaluation about Mouse

Compared to the 16% of positive evaluations about Mouse that were strengthened through GRADUATION or ENGAGEMENT, (12%) of the negative evaluations were strengthened. Again, the difference of this surface analysis is not statistically significant but it does not take into account the other strategies available to speakers to strengthen and weaken evaluations. The impact of these is discussed below in section 7.2.4.

	Composition Balance (N=104)		Composition Complexity (N=20)		Reaction Impact (N=4)		Reaction Quality (N=52)		Valuation (N=68)		Judgement (N=10)		Affect (N=32)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Force (Heighten)	10	10	0	0	0	0	2	4	2	3	2	20	6	19
Focus (Sharpen)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Engagement (Increase)	2	2	0	0	0	0	4	8	2	3	0	0	4	13
No Strengthening Strategy	92	89	20	100	4	100	46	89	64	94	8	80	24	75

Table 19 Distribution of different strengthening strategies for negative evaluations about Mouse

Table 21 shows that the distribution of strengthening strategies was quite unequal across the different categories of the Appraisal Theory Framework. The rows show the different strengthening strategies that were used and the columns show the different categories. In total there were 34 negative evaluations that were strengthened through the use of GRADUATION and ENGAGEMENT. Nearly two thirds of these contained COMPOSITION: BALANCE or AFFECT. There were no strengthened negative evaluations about the COMPOSITION: COMPLEXITY or REACTION: IMPACT. One interpretation of this is that the stimulus Mouse had a very strong impact at the first moment of truth and that this compensated for any problems with the actual design or dimensions of the pack. In line with this the category with the next fewest strengthened negative evaluations was VALUATION with 6%.

Examples of INTENSITY: HEIGHTEN are:

- 296. *I don't like that **at all**.*
- 297. *It's a bit the shape of it is **very** clumsy.*
- 298. *That's **completely** different isn't it.*

Examples of ENGAGEMENT: INCREASE are:

- 299. *Yea **definitely**.*
- 300. ***The fact** that it's plain makes it look like an air freshener.*

Chapter 6 has displayed the data in relation to the way in which evaluations about both stimuli were weakened and strengthened through the use of FORCE, FOCUS and ENGAGEMENT. Through the use of tables and examples from the transcripts I have provided data to help answer the questions that were raised in the introductions to sections 6.1 and 6.2. These issues and questions will be discussed in more detail in sections 7.1, 7.2 and 7.3 of the Discussion Chapter. The following section will explore the relationship between the use of weakening and strengthening strategies and an evaluations position in the turn taking structure.

6.3 Turn types in relation to weakening and strengthening strategies

As noted in section 5.4.2, the position that a turn takes in relation to previous turns can have a strong impact on the way in which an evaluation is realised. It can also help determine the level of strength that might be interpreted as existing within it. In terms of Face, disagreeing with the preceding speaker has a higher cost than an Initiation turn, particularly if there are significant power differences. However Initiation turns are less likely to have been influenced by the moderator or other participant as they occur when a speaker is either starting off a new topic or opening up a new dialogue. This means that the influence of peer pressure, leading questions or (dis)preferred responses is smaller in relation to other turn types. In conjunction with Fishbein and Ajzen's (1980) assertion that a speaker's initial assessments are the most salient and important, this might suggest that initiation turns inherently have a higher level of evaluative strength. In addition to looking at the weakening and strengthening strategies that have been used in relation to the Turn position, it will also be interesting to compare the initial evaluations to both the overall evaluations as measured by Appraisal Theory and also the score given in the final ranking exercise.

6.3.1 Initiating and Dispreferred turns in relation to Positive Evaluations about Egypt

Out of the 132 positive evaluations made about Egypt, 13.6% (18) were made within initiating turns. This could be seen as indicative of a stronger underlying evaluation but for many of these evaluations, the speaker also made use of different strategies to weaken the strength of their evaluation.

Turn Type	Egypt	%
Initiation (1)	18	13.6
RMNCD ¹ (2)	0	0
RMCD (3)	0	0
RPNCD (4)	0	0
RPCD (5)	0	0
RMC (6)	68	51.5
RPCA (7)	12	9.1
RPNCA (8)	6	4.6
RMCA (9)	2	1.5
RMNCA (10)	3	2.3
Total	109	

Table 20 Distribution of Turn Types for Positive Evaluations about Egypt

Table 22 shows that the 132 evaluations were made within 109 separate participant turns. In total there were 1066 different participant turns. This means that on average there was a positive evaluation approximately every ten turns. However where a turn contains evaluative content, there is likely to be more than one evaluation in addition to strengthening or weakening strategies that are being employed. For this reason the total number of evaluations is greater than the number of 'evaluative turns. This is important as it means that a moderator is likely to have to remember and interpret multiple evaluations and gradations at the same time.

Examples of FOCUS: SOFTEN are:

- 301. *That would be ok for like just a normal one to keep at home.*
- 302. *Modern it's sort of more.*

Examples of FORCE: LOWER are:

- 303. *I quite like to see that.*
- 304. *I quite like the er the groove thing.*

¹ The full text of these abbreviations is shown in section 2.1.3 Proposing a Scale of Inherent Strength Based on Structural and Social Factors.

Examples of ENGAGEMENT: DECREASE are:

- 305. *I think that's the best probably out of the three.*
- 306. *Bit easier I suppose with the ridges at the side for your fingers.*

In total there were seven evaluation turns of this type that did not make use of these weakening strategies. Of these remaining seven, two end in tag questions which as discussed above in section 2.1.1.5, can have a similar function as ENGAGEMENT:

DECREASE OR FORCE: LOWER:

- 307. *So that would be the spray then is that right?*
- 308. *It's to tell you where to put your hands isn't it?*

One is a conditional positive evaluation:

- 309. *I like the idea if it were smaller.*

The use of the subjunctive shows that the speaker only likes the idea in a hypothetical manner, dependent on certain changes being made to it. This means that there are four remaining positive evaluations that were not weakened in any way:

- 310. *I do like the idea of just a press down top.*
- 311. *You'd get a lot in it one bonus.*
- 312. *That'd be ok for home because it's a nice size and it stands up.*
- 313. *Contemporary yea.*

It is interesting to note that of these, none make use of any of the possible strategies for strengthening an evaluation.

As argued in section 2.1.3 there are 5 turn types that are seen as inherently strengthening an evaluation though for many of these there are no examples from my data. The turn type, of which there are examples in the data, which has the next highest inherent cost is a disagreement content turn with another participant. However these turn types were infrequent and in all of the exchanges about Egypt, there are only four where the participants explicitly disagree with each other. In each of these exchanges, the disagreement turn is a positive evaluation:

314. Speaker A: *No I don't like that at all I'd give that a wide berth looks cheap and.*

Speaker B: *I don't know there's something.*

Here it is clear that Speaker A's evaluation is negative and indeed the use of *at all* and *wide berth* act to strengthen its FORCE. Speaker B is less clear and directly disagrees, however to soften the blow to Speaker A's Face they weaken the disagreement by decreasing their level of ENGAGEMENT '*I don't know*'.

In the second example of positive disagreement turns, it is spread across a long exchange including a prompt from the moderator.

315. Speaker A: *Erm I guess the opposite to fresh or cool something that's more just dull just average just plain if you like nothing that you'd smell and think wow god I like that or what's that like.*

Moderator: *Yea.*

Speaker A: *Something that would really inspire me.*

Moderator: *Ok ok how about yourself.*

Speaker B: *No I think it would be cool.*

Speaker A: *Do you.*

Speaker B: *Yea I do.*

Moderator: *Ok.*

Speaker B: I don't know why I just think it would be cool for when you go to the gym.

All three of Speaker B's turns in this exchange are dispreferred responses. The first turn is in response to the neutral prompt from the moderator but it is clearly at odds with Speaker A's preceding evaluation on the same subject. The second turn has little content and is in direct response to Speaker B's request for clarification and the third turn is a summary that again explicitly disagrees with Speaker B's initial evaluation. In both of the full content turns Speaker B weakens the strength of their positive dispreferred evaluation with a decrease of ENGAGEMENT; *I think* and *I don't know*. Again, there are no instances where the positive, dispreferred response has been strengthened through the use of FORCE, FOCUS or ENGAGEMENT. This means that while the inherent cost of the evaluation is high, due to its place in the turn taking structure, the language that has been used has acted to counterbalance this inherent cost by weakening, rather than strengthening the evaluation. However the extent to which this balancing act is conscious or subconscious is up for debate and goes beyond the scope of this research.

6.3.2 Initiating and Dispreferred turns in relation to Negative Evaluations about Egypt

Of the 377 negative evaluations that were made about Egypt, 72 (19%) were made within initiation turns. While this figure is not significantly different from the 13.6% of positive evaluations about Egypt which were made within initiation turns, a closer analysis reveals that there were still differences.

Turn Type	Egypt	%
Initiation (1)	72	19.1
RMNCD (2)	0	0
RMCD (3)	0	0
RPNCD (4)	0	0
RPCD (5)	0	0
RMC (6)	177	47.1
RPCA (7)	39	10.4
RPNCA (8)	26	6.9
RMCA (9)	0	0
RMNCA (10)	4	1.1
Total	318	

Table 21 Distribution of Turn Types for Negative Evaluations about Egypt

Indeed while 75% of positive evaluations were weakened in some way, and none of the evaluations were strengthened, with negative evaluations it was quite different. Only 50% of the evaluations were weakened and this is statistically significant:

Examples of FOCUS: SOFTEN are:

- 316. *It's like a value pack.*
- 317. *I wouldn't pick that sort of it's too big.*

Examples of FORCE: LOWER are:

- 318. *It's a bit big for your bag.*
- 319. *I think that's quite bulky.*

Examples of ENGAGEMENT: DECREASE are:

- 320. *I guess it does say value pack.*
- 321. *I'm not sure I don't think I'd want to be carrying it around with me.*

In addition to the lower frequency of evaluations that were weakened, some of the negative initiation evaluations were actually strengthened.

Examples of FORCE: HEIGHTEN are:

- 322. *It's not nice **at all**.*
- 323. *I'd give that a **wide berth**.*

Examples of ENGAGEMENT: INCREASE are:

- 324. ***Definitely** wouldn't buy it.*
- 325. *I'd say it were **too big**.*

In total there were eight negative initiation evaluations that were strengthened in some way. This is in direct contrast to the positive evaluations of the same type where there were no examples that were both inherently strengthened by their position in the turn taking structure and simultaneously strengthened by the actual content of the turns.

Again, the turn type that has the next highest inherent cost is a disagreement content turn with another participant. There were two examples of this types that were negative evaluations.

- 326. Participant A: *It wouldn't stop me purchasing it.*
Moderator: *Yea.*
Participant A: *Then I would you know what I mean.*
Moderator: *Yea.*
Participant A: *I would then.*
Moderator: *Ok you seem a little less certain.*
Speaker B: ***I'm not sure I think it looks like a big cocoon as it is now or something you could play rugby with.***
Moderator: *Ok.*

In this first exchange Participant A gives their evaluation before the moderator prompts Participant B to give their own evaluation. In this instance the preferred response for Participant B to the Moderator would be to agree that they are a little less certain than Participant A. While this could be seen as a somewhat leading question, it is likely that the moderator suggested that they were a little less certain in response to some non-verbal cues that they were providing such as their body language or facial expressions. Speaker B does then take their turn which disagrees with Speaker A's evaluation. However within this turn Speaker B decreases their ENGAGEMENT – *I'm not sure I think* and softens the FOCUS – *It looks like, or something*, both of which act to lower the strength of the turn. The initial turn by Participant A is positive but the disagreement turn is in contrast to this and is a negative evaluation.

The second exchange involving disagreement is more straightforward as the moderator does not take a turn:

327. Participant A: *I'm not sure could be young young kids maybe eighteens nineteens maybe.*

Participant B: *I'd probably go I'd probably say older I'd think.*

By looking at the surrounding evaluations it is clear that Participant A's turn is meant as a negative evaluation. However, while Participant B disagrees with their assessment, it is unclear if this difference is enough to suggest that Participant B's evaluation is meant to be positive or if it is still negative. Again weakening strategies are used in the disagreement turn to downplay its strength- ENGAGEMENT: DECREASE *I'd probably go I'd probably say* and *I think*. In conjunction with the disagreement turns that were positive evaluations, there were no disagreement turns that were strengthened, or even left neutral and not weakened.

6.3.3 Initiating and Dispreferred turns in relation to Positive Evaluations about Mouse

Out of the 474 positive evaluations made about Mouse, 17% (82) were initiating turns. In comparison to the 14% of positive initiation turns about that were contained within initiating turns this difference is not statistically different but a closer analysis does show that there are still differences.

Turn Type	Egypt	%
Initiation (1)	82	17.3
RMNCD (2)	0	0
RMCD (3)	0	0
RPNCD (4)	0	0
RPCD (5)	6	1.3
RMC (6)	178	37.6
RPCA (7)	54	11.4
RPNCA (8)	42	8.9
RMCA (9)	10	2.1
RMNCA (10)	12	2.5
Total	318	

Table 22 Distribution of Turn Types for Positive Evaluations about Mouse

Indeed, while the majority of evaluations of this type were weakened for Egypt, it was quite different for Mouse. Although 33% of these evaluations were weakened using FORCE, FOCUS and ENGAGEMENT, the majority, 67% were not.

Examples of FOCUS: SOFTEN are:

- 328. *It looks like a heart almost when you at the right angle.*
- 329. *It looks more of a like Armani type.*

Examples of FORCE: LOWER are:

- 330. *I quite like the different button on it.*
- 331. *That's like quite like a oh that's a bit trendy.*

Examples of ENGAGEMENT: DECREASE are:

- 332. *Probably good for a man to put in his bag for the gym.*
- 333. *It's intriguing I think.*

Whereas none of the positive evaluations that were initiation turns for Egypt were strengthened, this was not the case for Mouse.

Examples of ENGAGEMENT: INCREASE are:

- 334. *I would definitely try that.*
- 335. *As long as the smell were nice I'd definitely try that.*

Examples of FORCE: HEIGHTEN are:

- 336. *It looks like you can get a good grip on it don't it.*
- 337. *It's quite easy enough to I fit my hand much better around that.*

In total there were seven evaluations that had their strength increased. However it is worth noting that of these seven, four of them also weakened the strength of the evaluation through the use of FORCE, FOCUS and ENGAGEMENT. There is also another example of a conditional evaluation:

- 338. *As long as the smell were nice I'd definitely try that.*

This leaves two remaining evaluations of this type that were strengthened inherently by their position in the turn taking structure and also by the content of what was said. As stated above, the turn type that has the next highest inherent cost is a disagreement content turn with another participant. However there were again very few exchanges of

this nature with only four taking place about Mouse, and of these only two had any evaluative content.

339. Participant A: *It's a bizarre shape for a deodorant isn't it because it sort of doesn't stand up or anything?*
Moderator: *No yea do you see that as being quite important for it to be quite a vertical thing.*
Participant A: *Er well I usually stand all my bits all together.*
Moderator: *Yea.*
Participant A: *Yea so probably be tidier.*
Participant B: *Suppose you could have it in your handbag.*
Participant A: *Bit big though isn't it.*

In this first example Participant A gives a negative evaluation on the basis that the pack cannot stand up. Participant B then disagrees that this has to be a negative attribute by pointing out that it could be used in a different way, for example, in a sports bag. Participant A disagrees with this though by saying it would still not be suitable due to its size. The initial disagreement turn is a positive one, disagreeing with the previous negative evaluation, and the subsequent disagreement turn is a negative evaluation, disagreeing with that positive evaluation. Both of the disagreement turns start off by making use of different strategies for downplaying the strength of their dispreferred response. Participant B decreases their ENGAGEMENT using *Suppose* and Participant A lowers their FORCE using *Bit*.

The second disagreement turn also had positive evaluative content:

340. Participant B: *No I don't like it.*
Participant A: *I quite like that cos it's different.*
Participant B: *I think that's more for a woman it would fit in your bathroom.*
Participant A: *Oh no I'd have thought it were for a man].*

Participant B: *Blend in*] would you I'd have said a woman.

Here the first speaker gives a clear and unmitigated negative evaluation which the second speaker disagrees with. Again this disagreement is prefaced by a weakening tactic as the FORCE is lowered through the use of *quite*. The second disagreement turn is a negative evaluation with the final disagreement turn being positive. In both of these instances the speaker makes a short utterance, *Oh no, would you*, which indicates disagreement, before decreasing their ENGAGEMENT, *I'd have thought* and *I'd have said*, prior to the main part of the turn that specifies what they are disagreeing with.

6.3.4 Initiating and Dispreferred turns in relation to Negative Evaluations about Mouse

Of the 291 negative evaluations that were made about Mouse, 66 (22.7%) were contained within initiation turns. Compared to the 17% of positive evaluations contained with initiation turns, this suggests that on the surface there is not a statistically significant difference between the way in which positive and negative evaluations about Mouse were strengthened through their position in the turn taking structure.

Turn Type	Mouse	%
Initiation (1)	66	22.7
RMNCD (2)	0	0
RMCD (3)	0	0
RPNCD (4)	0	0
RPCD (5)	4	1.4
RMC (6)	134	46
RPCA (7)	24	8.2
RPNCA (8)	8	2.8
RMCA (9)	10	3.4
RMNCA (10)	2	0.7
Total	248	

Table 23 Distribution of Turn Types for Negative Evaluations about Mouse

This is perhaps surprising given that Mouse was the most popular of the packs and that as noted previously, initial assessments can be seen as more indicative of overall evaluation.

However, as in previous sections, a closer analysis does reveal meaningful differences. Thirty two of these initiation turns that contain negative evaluations also contained weakening strategies that act to neutralise the inherent strengthening of the evaluation due to its place in the turn-taking structure.

Examples of ENGAGEMENT: DECREASE are:

- 341. *I probably wouldn't get that.*
- 342. *I think no unless I found out second hand it were good.*

Examples of FORCE: LOWER are:

- 343. *Interesting shape and I don't quite know what this is about.*
- 344. *I think it would appeal more if it was slightly smaller actually.*

Examples of FOCUS: SOFTEN are:

- 345. *I always worry with things like this.*
- 346. *Reminds me of an air-freshener.*

Several of these evaluations contained more than one weakening strategy or multiple examples of the same strategy:

- 347. *I think first impression is I don't like but who knows really I suppose.*
- 348. *looks like a bar of soap don't it?*

The last example also ends in a tag question which as noted in section 2.1.1.5 can be indicative of a lack of certainty and may therefore be considered analogous with ENGAGEMENT: DECREASE. Six of the negative evaluations, contained within an initiation turn, also contained positive evaluations:

349. *I think if you put it on the market people would buy it just to try it but I don't think that it would be a success.*
350. *Interesting shape and I don't quite know what this is about.*

It is interesting to note that in both of these instances the positive evaluation is made first with the negative evaluation being made afterwards. This means that out of the initial 66 negative evaluations that were strengthened through their position in the turn taking structure, only 30 were void of either positive evaluations or weakening strategies. Therefore, taking other factors into account means that only 10.3% of negative evaluations were contained within initiation turn and this is significantly lower than the 17% of positive evaluations which were contained within initiating turns.

For negative evaluations about Egypt, made within initiation turns, 11% were strengthened not only by their position in the turn takings but also through the use of graduation and engagement. However for evaluations about Mouse, made within initiation turns, there was only two such examples:

351. *I don't like that at all.*
352. *I really don't like that.*

There were two examples of disagreement content turns that had positive evaluations. However these were part of the exchange that included the disagreement content turns containing negative evaluations. These were discussed in section 6.3.3 above.

This chapter set out to present the necessary data to inform the discussion in the subsequent Discussion Chapter. It has also provided some preliminary analysis to help answer the questions set out in sections 6.1 and 6.2.

7. Discussion

The previous chapter set out to present the data with a view to answering several key questions in relation to the overall research questions as set out in Chapter 1. This Chapter will now attempt to answer those questions more fully and discuss these central points in more detail. This will be done by using examples from the transcripts to exemplify points in relation to approaches and theories that were introduced and explored in the literature review (Chapter 2). In section 6.1 I posed the following questions:

1. Is there a link between the overall ranking of the stimuli and the distribution of positive and negative evaluations?
2. Are different categories of the Appraisal Theory Framework more likely to attract positive or negative evaluations and does this have consequences with regards to the participants' overall evaluation of the pack?

Although the subsequent sections of Chapter 6 went some way towards answering these questions by describing the relevant data, there was little by the way of discussion of the data in relation to the literature review and this will be carried out below in section 7.1.

Once these areas have been discussed in more detail, section 7.2 will then look to answer the following question posed in the introduction to section 6.2:

3. Is there a link between the polarity of an evaluation and the distribution of strengthening or weakening markers? (I.e. are negative evaluations more likely to be weakened because of, say, politeness reasons?)

Section 7.3 will then explore the final question raised in section 6.2:

4. Is there a link between the distribution of strengthening or weakening markers and the explicit ratings given in the ranking exercise? (I.e. were participants

more likely to strengthen negative evaluations of the least liked pack or strengthen positive evaluations of the most liked pack?)

Again these questions were partially addressed throughout section 6.2 but will be explored in relation to the literature review in more detail below in section 7.2.

7.1 Polarity In Relation to Ranking and Category

As stated in the introduction to this chapter, this section will explore, in more detail, the questions set out in section 6.1. Initially this will involve exploring any link between the overall ranking of the stimuli and the distribution of positive and negative evaluations. Although I have argued throughout this thesis that the extent to which an evaluation has been strengthened or weakened is of high importance, this section will not account for these variations as they will be discussed below in section 7.2. The second half of this section will move on to explore the way in which different categories of the Appraisal Theory Framework might be more likely to attract positive or negative evaluations. As noted in section 6, Egypt was the least liked and Mouse the most liked from the explicit ranking exercises. There were 132 positive evaluations for Egypt compared to 474 for Mouse. Perhaps the most striking observation is the overall difference in the number of positive evaluations that were made with regards to the two stimuli. There were just over 3.5 times more positive evaluations made about Mouse than there were about Egypt. Mouse was rated as the most liked stimulus in the ranking exercises, and in terms of frequency, this is supported by the distribution of the positive evaluations. With regards to negative evaluations there were 377 for Egypt compared to 291 for Mouse. This means that there were 30% more negative evaluations for the least liked pack than for the most liked. While this difference is less striking than the difference in frequency of positive evaluations between the two packs, it is still statistically significant and suggestive of a link between the explicit rankings and the frequency and polarity of the less explicit linguistic evaluations.

One explanation for this difference in frequency could be the amount of time that the participants spent talking about each of the different stimuli. However as noted in section 6.3.3, the average time that each stimulus was discussed for was remarkably similar. For Mouse the average time was 8 minutes and 4 seconds and for Egypt it was 8 minutes and 12 seconds. In section 6.3.3 I raised the question of whether people have more to say when they feel strongly about something, regardless of polarity, or whether they say more about things that they like. Fishbein and Ajzen's (1980) research suggested that participants found it easier to talk about things that were important to them but did not distinguish between positive and negative evaluations. The data from the pilot study suggested that it was the strength of opinion, rather than polarity that was most closely correlated to the volume of evaluations that were elicited. In addition to this, where the participants were talking about a stimulus that they either strongly liked or strongly disliked, they were more likely to use explicit rather than evoked evaluations. This also ties in with Sivacek and Crano's (1982) work which suggests that the more interested the speaker is in the object that they are evaluating, the more likely they are to provide stronger opinions. However, although the total time taken to discuss each stimulus was similar, the number of evaluations that were elicited during these periods was less similar. In total there were 509 evaluations made with regards to Egypt compared to 765 that were made with regards to Mouse. Again this is statistically significant and means that there were 50% more evaluations for the most liked pack than the least liked. This would seem to suggest that the participants produced more evaluations in relation to their preference levels and not just their strength of feeling. However, it would perhaps be incorrect to suggest, on the basis of my data, that in general, participants make more evaluations about stimuli that they like, than stimuli that they do not like. One reason why there may have been more positive evaluations, even though the overall time spent talking about the stimuli was similar, could be related to issues of politeness and face. As discussed in section 2.1.2, research by Brown and Levinson (1987), Pomerantz, (1984) and White (1998) all suggests that when making a dispreferred response, the speaker is likely to use certain strategies to delay making their evaluation, and therefore the threat to face. Therefore, it would be expected that a length of discourse with more dispreferred responses would, in total, have more turns, and fewer containing evaluations, than one

with predominantly preferred responses. In section 2.1.1.3.1 I argued that a preferred response has a lower cost to the speaker than a dispreferred response and is therefore usually immediate and unmarked. Again this supports the idea that it might take longer to give dispreferred responses. Given that negative evaluations were more likely to be dispreferred responses, it follows that an hour's worth of talk containing predominantly negative, dispreferred responses, will contain fewer evaluations overall than an hour's worth of talk containing predominantly positive, preferred responses. Although I suggested in section 5.2.4 that each participant was encouraged to talk about each stimulus for as long as they were able, there were certain constraints. Each interview was only scheduled to last an hour and it was necessary to ensure that all of the topics and tasks on the discussion guide were covered. With more time, it may have proved that the participants would have made a higher number of negative evaluations, evening up the imbalance with the total number of positive evaluations that were made. Establishing whether, or not participants are likely to make more evaluations about stimuli that they like, than stimuli that they do not like could perhaps be an area for further study.

The frequency of negative evaluations about the least liked stimulus was statistically similar to the number of positive evaluations about the most liked stimulus. There were 377 negative evaluations about Egypt compared to 474 positive evaluations about Mouse. Although this means that there were 30% more negative evaluations about Egypt than positive evaluations about Mouse, compared to the number of positive evaluations for the least liked stimulus in relation to the number of negative evaluations about the most liked stimulus there is a significant difference. While there were only 132 positive evaluations with regards to Egypt, there were 291 negative evaluations given towards Mouse. This means that there were twice as many negative evaluations about the most liked pack than there were positive evaluations about the least liked pack and again this difference is statistically meaningful.

As suggested in the discussion on dispreferred seconds, the participants might have felt more able to make negative evaluations if they had already made some prior positive evaluations. Although providing a useful guide, simply analysing the frequency and

distribution of positive and negative evaluations does not make full use of the framework that has been developed within this research. Exploring the way in which these evaluations have been weakened or strengthened will also help establish the extent to which the scores from the ranking exercise compare to the less explicit evaluations from the interviews and will be looked at below in section 7.2. The second half of this section will now look at the relationship between different part of the Appraisal Theory Framework and their affect on the overall ranking of the stimulus.

As displayed in section 6.1.4, the distribution of negative evaluations regarding the COMPOSITION of the two packs was statistically similar. Negative evaluations containing COMPOSITION made up 46.1% of the total negative evaluations for Egypt and 44.3% of the total negative evaluations for Mouse. There were also a higher percentage of evaluations, with regards to positive evaluations containing COMPOSITION, for Egypt, 20%, than there were for Mouse, 14%. If the participants were ranking the packs purely on their ease of use (COMPOSITION: COMPLEXITY) and shape (COMPOSITION: BALANCE) then it would be expected that they would have scored similarly. The fact that Mouse and Egypt were the most and least liked pack suggests that the decision was based on additional factors. If this is the case then the distribution of other parts of the framework might give an indication as to why Mouse was preferred to Egypt. When the evaluations are looked at as a whole, rather than by polarity, the reasons for the difference in the ranking scores for Egypt and Mouse becomes clearer.

Egypt

	Positive	%	Negative	%	Total %
Composition	27	5.4	170	33.7	39
Reaction	51	10.1	108	21.4	31.5
Valuation	44	8.7	48	9.5	18.2
Affect	10	2.0	29	5.7	7.7
Judgement	4	0.8	14	2.8	3.6
Total					505

Table 24 Distribution of Positive and Negative Evaluations about Egypt across the Appraisal Theory Framework

Table 26 shows that for Egypt, a third of the evaluations were criticising the ease of use and shape of the stimulus. In addition to this a further fifth of the evaluations were

negative with regards to the level of QUALITY and IMPACT the stimulus had. This means that evaluations in these two categories make up just over half of the total evaluations. Combined with the very low level of positive evaluations with regards to JUDGEMENT or AFFECT, this suggests that the problems with the shape and design of the stimulus were sufficient to stop the participants liking the pack overall.

Mouse

	Positive	%	Negative	%	Total %
Composition	68	9	124	16.4	25.3
Reaction	152	20.1	56	7.4	27.4
Valuation	194	25.6	66	8.7	34.3
Affect	56	7.4	24	3.2	10.6
Judgement	8	1.1	10	1.3	2.4
Total					758

Table 25 Distribution of Positive and Negative Evaluations about Mouse across the Appraisal Theory Framework

While over 40% of the negative evaluations made about Mouse were with regards to its shape and ease of use, these made up a much smaller amount of the total number of evaluations, at just 16.4%. There were only half as many positive evaluations with regards to the COMPOSITION of the pack but this did not affect the positive evaluations in other parts of the Appraisal framework. Positive evaluations praising the QUALITY and IMPACT of the stimulus accounted for over a fifth of the total number and combined with positive comments about the overall VALUATION of the stimulus, this accounted for nearly half of the overall total. This suggests that despite the problems with the physical dimensions of the pack, the participants still liked it due to the overall design. Although evaluations with regards to COMPOSITION, IMPACT and QUALITY indicate how the participants felt about specific parts of the stimuli, evaluations relating to VALUATION and AFFECT are perhaps more indicative of how much a participant likes a stimulus. VALUATION, as stated in section 3.3.1.3, relates to the speaker's overall view and AFFECT relates to their emotional disposition to the stimuli. The fact that positive evaluations containing VALUATION was the most frequent category (25.6%), for the pack that received the highest scores in the explicit ranking exercise, supports this view.

This higher overall level of preference is also supported by the comparative totals for POSITIVE and NEGATIVE AFFECT which were significantly different. Only 2% of evaluations relating to Egypt had POSITIVE AFFECT and 8% NEGATIVE AFFECT. As stated in section 3.3.1.1 evaluations tagged for AFFECT indicate the speaker's emotional disposition to the stimulus. Where VALUATION perhaps captures the more rational reasons why a speaker might like the stimulus, AFFECT might be seen as accounting for those instances where the speaker cannot articulate what it is that they like, they just like it. As discussed above, although AFFECT is not graded in discrete categories, it can still be seen as consisting of lexical items that fit onto a scale, 'like', 'love' and 'adore', for instance. In all but one of the 2% of positive evaluations about Egypt that contained POSITIVE AFFECT, the lexical indicator was either 'like', which is at the bottom of the scale, or it was a non-content agreement turn which, I argue, is inherently the weakest possible form.

353. *I like the idea if it were smaller.*

354. *I quite like to see that.*

The only example that differed from this was:

355. *Aesthetically it looks quite nice pleasing.*

This suggests that there was no unconscious, 'I just like it' factor with this pack, to compete with all of the more rational negative evaluations that were made about it.

This contrasts with evaluations made about Mouse of which 8% had POSITIVE AFFECT and 3% had NEGATIVE AFFECT which is a significant difference. While Mouse was the most liked pack, the participants still made negative evaluations about it. As suggested below, the majority of the positive evaluations were made with regards to the IMPACT and QUALITY of the stimulus. There were a similar number of positive valuations containing AFFECT (7.4%) as there were with regards to the pack's COMPOSITION (9%). That is, the frequency of rational concrete evaluations and those which fit into the 'I just liked it'

category was statistically similar. For both Mouse and Egypt, the majority of the evaluations were tagged under the COMPOSITION, REACTION and VALUATION branches. Evaluations relating to AFFECT and JUDGEMENT made up just 11.3% of the total for Egypt and 13% for Mouse.

Only having two stimuli to base any correlations on makes it difficult to try and extrapolate out as to whether or not certain categories of the framework are more or less likely to have an effect on the overall preference levels as indicated by the explicit ranking packs. As suggested above, the VALUATION branch might best be compared to this explicit ranking as it covers utterances which assess the stimulus as an overall entity. It might be expected that Mouse would have a higher frequency of positive, and lower frequency of negative evaluations of this kind, and that Egypt would have a higher frequency of negative evaluations and lower frequency of positive evaluations of this kind. Table 14 shows that for Mouse, POSITIVE VALUATION was the highest frequency of any of the tags at 25.6%, and NEGATIVE VALUATION was fifth lowest out of the six main categories at 8.7%. A third of all evaluations made about Mouse were tagged under the VALUATION branch of the framework. For Egypt, evaluations tagged as positive VALUATION were the fifth least frequent out of the six main categories at 8.7%. This again supports the notion, as already suggested above, that the frequency and polarity of utterances tagged for VALUATION, might give a good indication of the overall preference levels. Overall, the same percentage of evaluations were tagged as positive VALUATION for the least liked pack as were tagged negative VALUATION for the most liked pack. This would suggest, however, that the frequency of NEGATIVE VALUATION for Egypt would be comparatively high. Perhaps surprisingly, Table 14 shows that this was not the case and there was statistical difference between negative evaluations, 9.5%, and positive evaluations, 8.7%, for this branch of the framework. One reason for this could be the high frequency of negative evaluations regarding COMPOSITION (33.7%) and REACTION (21.4%). These totaled more than half of all the evaluations about this stimulus, so the participants might have felt that their opinion of the stimulus was already clear. They might then have felt that they did not need to reiterate it with the kind of summary evaluations that would be tagged under the VALUATION branch of the framework. Also, as

discussed above, negative evaluations were likely to take longer to make, meaning that they were less frequent in general, than positive evaluations. In addition to this, there were some instances where I had to move an interview on to the next stimulus, to make sure that the interviews did not overrun. It therefore, might be reasonable to assume that the evaluations that were missed were more likely to be the kind of summary evaluations that would be tagged under the VALUATION branch of the framework.

7.2 *Weakening and Strengthening Strategies in Relation to Polarity*

This section will explore, in more detail, the questions set out in section 6.2. It will explore any link between the polarity of an evaluation and the way in which it has been weakened or strengthened. To assess whether there is any link between the polarity of an evaluation and weakening or strengthening features that have been used, I will first look at positive evaluations followed by negative evaluations. Within these sections, these questions will be further explored taking the individual stimuli into account. Although posed as separate questions in the introduction to this thesis, it has become apparent that any inherent strengthening or weakening of an evaluation is intrinsically related to the social and interactional context within which it is made. The weakening or strengthening of the evaluations is indicated by both the evaluation's position in the turn taking structure and also the lexis that is used within that turn. Therefore the first two research questions are both dealt with in this section.

7.2.1 Weakening of Positive Evaluations

Of the 132 positive evaluations that were made about Egypt, 58% (77) used one or more of the possible weakening strategies to downplay the strength of their positive evaluation. The use of these different strategies could be a result of two main factors. Firstly it could be that the speaker did have something positive to say about the stimulus but it was not a strongly held conviction and so the language that they used indicates this

by hedging the opinion. The alternative is that they did have a stronger positive opinion about the stimulus but perceived power differences, social norms, and politeness conventions meant that they felt that they had to downplay their strength of opinion.

Of the 474 positive evaluations about Mouse, 210 (44.3%) were weakened using one or more of the available strategies from GRADUATION or ENGAGEMENT. The distribution of positive evaluations that were weakened was quite even across each of the different parts of the framework, varying from 45% to 58% of the evaluations being weakened. Only COMPOSITION: BALANCE (12%) and REACTION: IMPACT (100%) significantly varied from this. In comparison to the same category of evaluations for Egypt, a statistically significant higher proportion remained unweakened, 57.1 % compared to 45.3%. As Mouse was the preferred stimulus from the final ranking exercise, this might suggest that the participants felt more strongly about their positive evaluations and therefore did not feel the need to weaken them as frequently. The distribution was also different across the two stimuli as for Egypt it ranged from 75% being weakened (REACTION: QUALITY) to only 25% being weakened, (COMPOSITION: BALANCE and COMPOSITION: COMPLEXITY); REACTION: QUALITY and VALUATION were the parts of the framework that had the most weakened positive evaluations.

In total there were 606 positive evaluations made about the two stimuli. Of these, 346 or 57% were weakened through FORCE: LOWER, FOCUS: SOFTEN and ENGAGEMENT: DECREASE. Many of these evaluations used a combination of these to further weaken the evaluation. It is interesting to note that, on the surface, overall preference levels were not statistically significant with regards to the ways in which positive evaluations were weakened, 44.3% for Mouse, and 58.3% for Egypt.

7.2.2 Strengthening of Positive Evaluations

Of the 132 positive evaluations that were made about Egypt only fourteen (10.6%) were strengthened through GRADUATION or ENGAGEMENT. Of these fourteen instances, twelve also included markers that weakened the evaluation.

Examples of ENGAGEMENT: DECREASE are:

- 356. *Definitely probably erm definitely a male product.*
- 357. *I would probably say that it would be more expensive.*
- 358. *Obviously for guys I think.*

Examples of FOCUS: SOFTEN are:

- 359. *It to me it's like a young design a trendy design.*
- 360. *Looking bordering on futuristic I would say.*

It is interesting to note that in all but one of these occurrences, the strengthening of the opinion comes first and is then followed by a lowering of strength. Though in number 362 there is then a subsequent strengthening strategy used.

- 361. *Obviously* (ENGAGEMENT: INCREASE) *for guys I think* (ENGAGEMENT: DECREASE).
- 362. *Definitely* (ENGAGEMENT: INCREASE) *probably* (ENGAGEMENT: DECREASE) *erm definitely* (ENGAGEMENT: INCREASE) *a male product.*

This may be because the speaker realises that they are expressing a strong opinion, but is not sure that they want to commit to it, and so subsequently attempts to soften it in some way. This means that there are only two positive evaluations that are genuinely strengthened.

- 363. *Yea definitely again and again yea.*
- 364. *Certainly wouldn't think it were cheap.*

As noted above in 5.4.2, the different role that a turn takes in a conversation also seems to affect the level of strength that should be associated with an evaluation so I will now look at the turn types that these examples come under. Example 365 comes from an exchange

with the moderator in relation to whether or not the participant thinks that they would buy the pack:

365. Participant: *I reckon that would be a seller.*
Moderator: *So I mean you would see yourself buying that?*
Participant: *Yea definitely again and again yea.*

The final turn is an agreement-content turn in response to a moderator question that is seeking to clarify an earlier response. Therefore it is a weaker evaluation than if the participant had simply stated that they would definitely buy the product on the basis of the stimulus in front of them. In their previous turn in this exchange the participant states that they think that the product would sell but they mitigate this with the use of '*I reckon*' and they also frame the assertion in a way that avoids any agency with regards to who would do the buying. So even though the final turn is a positive evaluation that has been strengthened, its position in the exchange structure means that it is still perhaps weaker than it might first seem.

The second example '*Certainly wouldn't think it were cheap*' is tagged as positive because being seen as cheap or low quality is a negative evaluation. By looking at the local context it is perhaps possible to establish that in this instance 'cheap' is seen as a negative rather than a positive evaluation and that an expensive looking product is more desirable.

366. Moderator: *Would you expect that to be a top end brand or lower or what do you think?*
Speaker A: *Yea mm I think so yea it's it certainly wouldn't think it were cheap.*
Speaker B: *Just I would middle to upper I would probably say.*
Speaker A: *yea it's not cheap you know don't think it would be a cheap looking one Tesco's own.*

While the participant has reiterated their evaluation that the product is not cheap, this is not the same as saying that it is premium or expensive looking. In some ways this could be considered a neutral evaluation as it is neither explicitly positive, '*I'd certainly say it were expensive*' nor explicitly negative '*I'd certainly say it were cheap*'. However as noted in the introduction to this Chapter, Appraisal Theory does not make much provision for such neutral evaluations. Although Martin and White (2005: 93) state that they are interested in whether people and their evaluations '*...present themselves as standing with, as standing against, as undecided, or as neutral with respect to these other speakers and their value positions*', Appraisal Theory does not seem to have any way of accounting for the polarity of an evaluation that is neither positive nor negative. Hunston and Thompson (2003: 3) state that opinions '*can be seen essentially in terms of positive and negative*'. This doesn't allow any room for a neutral evaluation. This notion of neutral evaluations is revisited and discussed in further detail in section 7.4.2.

So, while initially it seemed that there were fourteen positive evaluations that were strengthened; a closer inspection has shown that there were only two that did not also include some weakening features and that even these were weakened by their role in the turn-taking structure or by a possible lack in the framework. This means that there are no positive evaluations about the stimulus Egypt that have been definitively strengthened.

The remaining 16% (22) of the positive evaluations were neither explicitly weakened nor strengthened. Nine of these occurrences were non-content turns simply agreeing with the previous participant's evaluation:

367. *I agree.*

368. *Yea.*

369. *Mmm yea.*

This means that there were thirteen positive evaluations that did not make use of the available options to strengthen or weaken their opinion. Having looked at the way in

which positive evaluations about Egypt were strengthened, this section will now do the same for Mouse.

In total, of the 474 positive evaluations about Mouse, 76 or one sixth (16%) were explicitly strengthened through the use of ENGAGEMENT: INCREASE or FORCE: HEIGHTEN. Positive evaluations that related to the IMPACT and the QUALITY of the pack and its overall VALUATION accounted for nearly three quarters of the total positive evaluations. There were no strengthening strategies in evaluations that related to COMPOSITION: BALANCE, COMPOSITION: COMPLEXITY, JUDGEMENT or AFFECT.

Examples of REACTION: IMPACT are:

- 370. *I'd look at it definitely.*
- 371. *Yea absolutely.*

Examples of REACTION: QUALITY are:

- 372. *Yea I'd say more expensive yea.*
- 373. *It's more decorative so for dressing tables I'd say.*

Examples of VALUATION are:

- 374. *I would I would certainly want to try that one.*
- 375. *It's so so novel.*

Approximately 20% of the evaluations containing REACTION or VALUATION were strengthened. Taking into account which evaluations have been strengthened is likely to be beneficial, when wanting to establish which categories are more important to the participants. Firstly, the explicit strengthening shows which evaluations they felt most strongly about. In addition, as discussed above, speakers are unlikely to strengthen evaluations about stimuli that overall, they do not like. While they may still make these

positive evaluations for reasons of politeness or face, or to try and take over the conversational floor, they are less likely to strengthen them. This also means, therefore, that giving more weight to those evaluations which have been strengthened will reduce the chance of placing too much importance on evaluations which were made due to social and interactional pressures.

The fact that there were more strengthened positive evaluations, about Mouse, with regards to REACTION and VALUATION, suggests that the stimulus had a strong effect at the first moment of truth. In conjunction with its likeability, this led the participants to give Mouse a positive evaluation overall. This correlates with the figures from section 6.1.2 which suggested that it was the good first impression that the pack made that led to the subsequent high overall ranking, rather than the actual design or ease of use or other factors more associated with the second moment of truth.

As noted above, many of the evaluations made use of both weakening and strengthening strategies within the same evaluative turn. These will now be discussed in relation to the weakening of positive evaluations for Mouse, before the different roles that the evaluative turns take in the exchange structure is analysed. Of the 80 strengthened positive evaluations, 28 also make use of strategies that act to weaken the evaluation. These include ENGAGEMENT: DECREASE, FOCUS: SOFTEN and FORCE: LOWER.

Examples of ENGAGEMENT: DECREASE are:

376. *I think more expensive because of the shape.*

377. *Yea I would probably try it to be honest.*

Examples of FOCUS: SOFTEN are:

378. *I'd say a pebble off a beach type thing.*

379. *I could say yea actually state of the art deodorant like you know twenty-first century deodorant.*

Examples of FORCE: LOWER are:

380. *It's a bit different and it looks completely different to what else is on the market.*
381. *It's quite easy enough to I fit my hands much better around that.*

Of the remaining 52 evaluations, nine of them were agreement non-content turns. This indicates that they have a weaker evaluative strength.

382. Moderator: *Ok erm so I mean you've said definitely female.*
Participant A: *Yes.*
Participant B: *Yea.*

Six were agreement content turns to the moderator, which is also a weaker evaluation.

383. Moderator: *So I mean would that be more expensive then.*
Participant A: *Yea I'd say more expensive yea.*

This leaves 22 which were content turns that were either in response to the other participant (384) or the moderator (385):

384. Participant A: *But it's quite easy enough to I fit my hands much better around that.*
Participant B: *Yea it is a lot better than that other one.*
385. Moderator: *Ok so if we had that on the shelf?*
Participant A: *I'd look at it definitely.*

There were also fifteen initiating turns, which, due to their position in the turn-taking structure, indicates that they were stronger evaluations.

386. Participant B: *I would definitely try that.*

Participant A: *It does look more trendy.*

This means that out of an initial 80 positive evaluations that were strengthened, only 37 were not weakened once other weakening markers and the evaluation's position in the turn taking structure had been considered.

On the surface, the difference between the number of positive evaluations that were strengthened for Egypt (10.2%) and Mouse (16.7%) is not statistically significant which is perhaps surprising given that these were the most and least liked packs. It might have been expected that the participants would make more strengthened positive evaluations about the most liked pack than they would about the least liked pack. However, as discussed above, a close analysis of the strengthened, positive evaluations for Egypt actually showed that there no evaluations that had not also been weakened in one of three different ways:

- 1) The use of weakening strategies from the Appraisal Theory Framework
- 2) The role it took in the turn-taking structure
- 3) Weakening lexis

In comparison there were 37 such evaluations for Mouse. It is therefore apparent that strengthened evaluations that were either neutral or strengthened in terms of their position in the exchange structure and that were free of weakening lexical markers were statistically more frequent for Mouse than for Egypt.

This is one of the main findings from this research. In general, I found that although the participants made positive evaluations about stimuli, that overall, they did not like, these evaluations rarely made use of the available strategies for strengthening an evaluation. Looking at how evaluations were strengthened through an Appraisal analysis, this became clearer. While initially it appeared that there were fourteen positive evaluations that had been strengthened, a closer analysis showed that all of these were also weakened

in some way. As indicated above, an initial analysis of the two stimuli would suggest that a similar frequency of positive evaluations were strengthened, regardless of the overall preference levels for each pack. However this similarity is shown, to some extent, to be false, when a more detailed analysis is carried out.

This is also supported by the results when analysing from a turn taking position. In total there were eighteen positive evaluations about Egypt that were initiating turns. In this thesis I have argued that an initiating turn is inherently strengthened due to its position in the turn taking structure. However of these eighteen, ten also made use of the weakening strategies of FORCE: LOWER, FOCUS: SOFTEN and ENGAGEMENT: DECREASE. In addition to this, two of the remaining positive evaluations ended with tag questions which as argued above, are comparable to ENGAGEMENT: DECREASE or FORCE: LOWER. This meant that where originally it appeared that there were eighteen positive evaluations about Egypt that were inherently strengthened, a closer analysis taking all of the available information into account, suggests that there were only six evaluations of this kind.

In total there were 606 positive evaluations made about the two stimuli. Of these, 94 or 15.5% were strengthened through FORCE: HEIGHTEN, FOCUS: SHARPEN or ENGAGEMENT: INCREASE. Many of these evaluations used a combination of these to further strengthen the evaluation. Again it is interesting to note that the participant's overall preference levels did not lead to statistically significant differences in the number of positive evaluations that were strengthened, 10.6% for Egypt and 16% for Mouse.

Having looked at the way in which positive evaluations were weakened or strengthened across the two stimuli, the following sections will do the same for negative evaluations.

7.2.3 Weakening of Negative Evaluations

Of the 377 negative evaluations about Egypt, 47% (177) made use of one or more of the possible weakening strategies to downplay the strength of the negative evaluation. This is statistically different to the 58% of positive evaluations that used this weakening

approach. In addition, a closer analysis shows further differences. Where negative evaluations were weakened, they tended to only use one weakening strategy:

- 387. *It's quite big.*
- 388. *That one's a bit of a waste there.*
- 389. *I don't think I'd make a purchase.*

In comparison, where positive evaluations were weakened they were more likely to contain more than one strategy or multiple occurrences of the same strategy:

- 390. *So I think that's quite useful.*
- 391. *It reminds me of a candle like erm a trendy candle or something.*
- 392. *I think that probably is a little bit more expensive looking.*

This means that although the total number of evaluations that were weakened was more similar than might be expected, the total number of weakening strategies used was much higher for positive evaluations than for negative evaluations.

This suggests that for this stimulus the speakers found it more difficult to make unmitigated positive evaluations than unmitigated negative evaluations. As discussed above in section 5.4.2, the strong trend towards agreement that the social, situational and interpersonal pressures cause, may be the reason for this. Although there was some variation, the general trend for this pack was that it was the least liked of all the stimuli, therefore if the other participant was making more negative evaluations, then the second speaker may have felt more comfortable doing the same. This in turn may have made them more likely to mitigate their differing, in this case positive, evaluations.

- 393. Participant A: *Erm I guess the opposite to fresh or cool something that's more just dull just average just plain if you like nothing that you'd smell and think wow god I like that or what's that like.*

Moderator: *Yea.*

Participant A: *Something that would really inspire me.*

Moderator: *Ok ok how about yourself.*

Participant B: *No **I think** it would be cool.*

While it has been established that there were very few incidences of direct disagreement, it is likely that participant would be aware of the prevailing opinion that the other speaker held with regards to the stimulus. This however does not show why the first speaker may have made the negative evaluations to start with. To try and gain some insight into this I will also explore these patterns in relation to the different turn-types that each evaluation was comprised of. This is discussed below in section 7.3.

Of the 291 negative evaluations about Mouse, 148 (50.9%) made use of one or more of the possible weakening strategies to downplay the strength of the negative evaluation. This is not significantly different to the percentage of positive and negative evaluations about Egypt that were weakened, 58% and 47% respectively, but is significantly different to the percentage of positive evaluations about Mouse that were weakened, 44.3%. The three categories that had the highest frequency of positive evaluations that had been strengthened, REACTION: IMPACT, REACTION: QUALITY and VALUATION, were also the three categories that had the highest frequency of weakening strategies for negative evaluations. This suggests that not only were these the areas that the participants felt most comfortable praising, they also wanted to mitigate any negative comments that they did make across these areas.

In terms of exploring the way in which all of the evaluations have been weakened, across both stimuli, the evaluation type that stands out is positive evaluations about Mouse. These evaluations were much less likely to have been weakened than any of the other evaluations. However it is apparent that across both stimuli and accounting for both positive and negative evaluations, approximately half of all evaluations are weakened in some way. This again is one of the key findings from the research. It is perhaps surprising that neither the polarity of the evaluations nor preference levels towards the pack had a significant impact on the percentage of evaluations that were weakened. For a

variety of reasons approximately half of all evaluations were weakened or mitigated in some way. One reason for this could simply be the design of the research methodology. The participants were asked to provide opinions on stimuli that they were unlikely to feel strongly about. In addition to this, they were aware that their opinions were being recorded and that their views were likely to be looked at once they had made them. Although I made every effort to reassure the participants, it is likely that this would still have had some impact on how confident they felt when giving opinions. In addition to this, the paired nature of the research design might also have had an impact on this aspect of the evaluations, as providing opinions in front of a stranger; who they knew was not there in a professional capacity; might also have led the participants to mitigate or hedge their opinions for fear of providing a 'wrong' or socially abnormal answer. Again, while I tried to ensure that the participants were relaxed and knew a little about each other before we started, it is possible that this would have some effect on their answers. It would be interesting to carry out research to ascertain what impact the environmental and interpersonal context has on the frequency of weakening or strengthening of evaluations.

7.2.4 Strengthening of Negative Evaluations

Of the 377 negative evaluations about Egypt, 67 (17.8%) made use of one or more of the possible strengthening strategies to increase the strength of the negative evaluation. 80% of the positive evaluations that increased their strength also contained some markers to weaken them. For negative evaluations the opposite was true as 84% of the evaluations that had been strengthened were free from markers that also weakened them in some way. On those occasions where they were also weakened, both GRADUATION and ENGAGEMENT were used:

Examples of FOCUS: SOFTEN are:

394. *Looks more like a shampoo bottle than a deodorant.*
395. *It just like screams out as though it should be called Aztec or something.*

Examples of FORCE: LOWER are:

- 396. *It's quite ugly ok I think personally.*
- 397. *I'd say it's a little bit just a bit too big.*

Examples of ENGAGEMENT: DECREASE are:

- 398. *I'd say if that was maybe a bit thinner that would help I think.*

It was shown above that the participants seemed to find it easier to make negative evaluations than positive ones and while there were no incidences of strengthened positive evaluations, there are many negative evaluations that have been strengthened. This would suggest that those participants who disliked the pack felt more strongly than those who indicated that they did like it.

From a power relationship perspective, it might be expected that the participants would find it easier to make positive evaluations than negative ones. Although it was made clear that, as the moderator I had no affiliation with either the individual stimuli or Unilever, it might still be expected that the participants would feel that they were making a face threatening act by being overly critical. It is likely that the participants presumed that I had a preference for the packs to be liked, and because I was carrying out the interviews, that would also be my expectation. This can be seen with utterances such as:

- 399. *I don't want to you know be all negative against these designs of yours.*
- 400. *Yea I'd say that was your best design of all of them.*
- 401. *Sorry for keep being negative about what you've done.*

Therefore making strong negative evaluations would clearly go against this expectation.

In comparison to the 16% of positive evaluations about Mouse that were strengthened through GRADUATION or ENGAGEMENT, a statistically similar number, 36 (12%), of the negative evaluations were also strengthened. As the most liked pack it would be

expected that there would be fewer examples of negative evaluations that had been explicitly strengthened. Of these 36 examples, however, many of them also included weakening strategies which act to neutralise the strengthening strategies.

Examples of ENGAGEMENT: DECREASE are:

- 402. *I don't think I'm overly impressed.*
- 403. *I'd probably think more expensive.*
- 404. *It's a bit the shape of it is very clumsy.*

Examples of FORCE: LOWER are:

- 405. *It's just really odd.*
- 406. *Erm just doesn't do anything for me at all.*

In total, 20% of the negative evaluations that were strengthened also contained weakening strategies. In section 7.2.2 I argued that one of the main findings from this research is that participants tend not to strengthen positive evaluations about stimuli that overall, they do not like. While initially it seemed that there were 32 positive evaluations about Egypt, that were strengthened, either semantically or through their position in the turn taking structure, a closer analysis showed that there were only six that were not also significantly weakened using the same strategies. I will now look in more detail at the way in which negative evaluations about Mouse were strengthened and compare this to the way in which positive evaluations about Egypt strengthened.

In section 6.3.4 I displayed the results with regards to the position in the turn taking structure, of negative evaluations about Mouse. Initially it appeared that nearly a fifth of these negative evaluations had taken place within Initiating turns. However, a closer analysis showed that more than half of these evaluations that were inherently strengthened through their position in the turn taking structure, also contained positive evaluations or weakening strategies of ENGAGEMENT: DECREASE, FORCE: LOWER or FOCUS:

SOFTEN. In this section I have shown that there were a surprisingly high number of negative evaluations, about Mouse, that had been strengthened, even when accounting for those that were also weakened in some way. However, when the analysis of the way in which the evaluations have been strengthened, is combined with the evaluation's position in the turn taking structure, a much clearer pattern emerges. Of the remaining thirty negative evaluations that had been strengthened through word choices that the participants made, only two of these were also inherently strengthened by the evaluation's position in the turn taking structure. This means that the remaining 28 negative evaluations about Mouse that had been strengthened, were made within turns that were inherently weaker. This might suggest that where the participants were unconsciously strengthening the negative evaluation through the use of initiating or dispreferred turns, they moderated this through their choice of words.

The main finding from this section, therefore, is that there were only two examples of a negative evaluation about Mouse that had been strengthened, both through its position in the turn taking structure, and through lexical strengthening strategies:

407. *I don't like that at all.*

408. *I really don't like that.*

This suggests that in addition to participants tending not to strengthen positive evaluations about stimuli that overall, they do not like, they also tend not to strengthen negative evaluations about stimuli that overall, they do like. Intuitively this seems to make sense but would perhaps not have been so obvious if the data had not been analysed with an approach combining Politeness, Turn Taking and Appraisal Theory analyses. One reason for the appearance of unstrengthened, positive evaluations, about stimuli that they did not like, could be for reasons of politeness and face. As demonstrated above, despite my best efforts, the participants, to some extent, still regarded the stimuli as 'belonging' to me. Therefore, if they were aware that they had been making lots of strengthened, negative evaluations, it is likely that they might have felt the need to counteract that by making some positive evaluations. However, because these positive

evaluations were being made for politeness reasons, rather than because the participants genuinely liked the feature that they were praising, they were much less likely to have been strengthened. This can be seen in the following exchanges:

409. Speaker A: *I mean it's ugly.*
 Speaker B: *Yea it is.*
 Moderator: *Ok.*
 Speaker A: *It's not nice at all.*
 Speaker B: *It's not even comfy I wouldn't have said not even for a man*
 Speaker A: *You'd get a lot in it one bonus.*

Here Speaker A has made two negative evaluations, one strengthened by its position in the turn taking structure and one through the use of FORCE: HEIGHTEN. However they then make a positive evaluation that acts to try and mitigate the strength of these previous negative evaluations. However, perhaps because it is produced for reasons of politeness rather than being an honest assessment, it remains unstrengthened.

In summary, I have argued that in terms of the lexis that was used, the polarity of the evaluation had little influence on the way in which evaluations were strengthened or weakened.

Stimulus	Egypt		Mouse	
	Positive	Negative	Positive	Negative
Weakened	58%	47%	44.3%	50.9%
Strengthened	10.6%	17.8%	16%	12%

Table 26 Frequency of Weakening and Strengthening Markers across both Stimuli

Table 28 shows that thon a surface level there was surprisingly little difference in the way which evaluations were weakened or strengthened, regarding the polarity of the evaluation, and also the stimulus that was under evaluation. However this section has argued that a more detailed analysis, accounting for the social and structural elements of

communication demonstrates more marked differences. Additionally, though it is difficult to make any strong claims on the basis of two stimuli, it is apparent that the most liked stimulus had fewer strengthened negative evaluations and more strengthened positive evaluations than the least liked stimulus.

7.3 *Weakening and Strengthening Strategies in Relation to Ranking*

Having looked at the link between the polarity of an evaluation and the way in which it has been weakened or strengthened, this section will now explore the way in which these weakening and strengthening markers relate to the explicit ratings given in the final ranking exercise. Although these issues have been partially discussed in section 6.2, this was segmented by the polarity of the evaluation and therefore a separate summary is provided below without this additional segmentation.

In total there were 509 evaluations containing attitude markers with regards to Egypt. Of these, 254 (49.9%) used one of more of the different approaches from ENGAGEMENT and GRADUATION to weaken the evaluation. The percentage of positive evaluations that were weakened (55%) was not significantly different to the number negative evaluations that were weakened (47%). Of the 509 evaluations about Egypt that were marked for attitude, 81 (16%) were strengthened through the use of GRADUATION or ENGAGEMENT. In comparison to evaluations regarding Egypt that were weakened, polarity had a stronger effect on evaluations about Egypt that were strengthened. Negative evaluations were significantly more likely to have been strengthened (17.8%) than positive evaluations (10.2%).

In total there were 765 evaluations containing attitude markers with regards to Mouse. Of these 358 (47%) used one or more of the different strategies from ENGAGEMENT and GRADUATION to weaken the evaluation. This is not statistically different to the percentage of evaluations about Egypt that had also been weakened (49%). Also similar to Egypt, there was no statistical difference between the number of positive evaluations

that had been weakened (44.3%) and the number of negative evaluations that were weakened (51%). While for Egypt it was slightly more likely that positive evaluations would be weakened, for Mouse, the opposite was true. Of the 765 evaluations about Mouse that were marked for attitude, 110 (14.4%) were strengthened through the use of GRADUATION or ENGAGEMENT. Again this is not statistically different to the percentage of evaluations about Egypt that were strengthened (16%). While for Mouse, there were 69% more positive evaluations (474) than negative evaluations (284), the percentage of positive evaluations that were strengthened (16%) was similar to the percentage of negative evaluations that had been strengthened (12%). Overall, the percentage of evaluations that were either weakened or strengthened was not statistically different across both stimuli; just fewer than 50% of evaluations marked for attitude were weakened, (Egypt 49%; Mouse 47%) and approximately a seventh of evaluations marked for attitude were weakened, (Egypt 16%; Mouse 14%). Having attempted to answer the questions that were first raised in Chapter 6, section 7.4 will then look at areas where I felt that the Appraisal Theory Framework did not fully account for the way in which evaluations were being realised and propose suggested modification to the Appraisal Theory framework.

7.4 Shortcomings and Suggested Modifications

I have argued within this thesis that a linguistic analysis allows the researcher to account for all of the evaluative content of a discourse and that this could be used to provide a different kind of insight than might be achieved through more standard market research approaches. Appraisal Theory has been used as the primary framework of analysis and has proved to be a robust and suitable approach. However, there have been occasions where I have felt that I have not been able to tag an evaluation in a way that fully captures what I believe the speaker to have meant. This has happened across two main areas; the concept of neutral evaluations and the relevance of the evaluation to the speaker. Sections 7.4.1 and 7.4.2 will explore these areas and suggest how the framework could perhaps be extended to account for such evaluations.

7.4.1 Engagement: Relevance

Although Appraisal Theory was not designed with a market research perspective in mind, it maps onto the areas of products, brands, or marketing communication that market researchers might be interested in. However, one area where the framework could perhaps be extended is with regards to Relevance; how relevant is the item or idea that is being evaluated, to the person making the evaluation? As it stands, this is perhaps best placed within the Engagement part of the framework. It was decided early on in the exploration of Appraisal Theory that it would not be possible or advantageous to use all of the Appraisal Theory Framework to the fullest levels. Therefore the Engagement branch was simplified to only show whether the level was increased or decreased. Where a heteroglossic evaluation was made, this was seen as an indication of lower evaluative force or strength as it allowed for alternative viewpoints. It therefore inherently contains an acceptance that the speaker's evaluation might be viewed as incorrect by other participants. Conversely, where a monoglossic evaluation was made, this was seen as indicative of a stronger evaluative force as it discounted any alternative viewpoints. Although this simplification of the Evaluation branch does miss out some subtleties in terms of the way in which the hetero- or monoglossic stance is framed, it is broadly in line with Martin and White's thinking. They state that:

'In broad terms, then, we can categorise utterances accordingly to this two-way distinction, classifying them as 'monoglossic' when they make no reference to other voices and viewpoints and as 'heteroglossic' when they do invoke or allow for dialogistic alternatives.' Martin and White (2005: 100)

I have reiterated this point here as one of the finer grade distinctions that Martin and White make within the Engagement branch is in terms of the Attribution of the evaluation. This distinction is made at the next level on from the Heteroglossic branch of the Appraisal Framework and as discussed in section 2.4.3.2, it is split into two further branches of Attribute: Distance and Attribute: Acknowledge. This notion of Attribution seems to be closely related to that of Relevance. Where a speaker feels that they are not

the right person to give an evaluation or that a different demographic is more relevant, they may project an evaluation onto that individual or group. Examples of this are shown below.

While the Attribute branch of the framework is suitable for reporting what other people have said or believe, this is very different to guessing at what their opinions would be. Due to the nature of this research methodology, it was impossible for participants to report the opinions of others. There were however, several instances of participants projecting their opinions onto others and again this seems to be slightly different to what the Attribute branch of the framework is appropriate for. These projections were onto individuals, specific, different demographic groups and more general 'others'.

Examples of projecting onto specific individuals are:

- a) *Johnny my boyfriend would definitely like that because it's more masculine.*
- b) *Certainly my husband wouldn't pick that up and think ooh yea I'll have a go at that.*

Examples of projecting onto a specific demographic group are:

- a) *Maybe it would appeal to older people.*
- b) *Probably men would like that because it's more sporty.*

Examples of projecting onto non-specific 'others' are:

- a) *Yea I probably wouldn't go for it myself but I think some people would find it attractive.*
- b) *I wouldn't buy it for me but there might be some people different to me.*

In each of these instances the speaker is not reporting back what their boyfriend or older people or some other people had said, instead they are attempting to guess at what they

might think. It is interesting to note that where the projection is on to a specific person, and one that the speaker knows well, there is a high level of certainty about the evaluation that they are projecting. In the two examples above the adverbs *definitely* and *certainly* are used. In contrast, when the projection is on to an unknown 'other' the level of Engagement is decreased through the use of *maybe*, *probably*, *I think* and *there might be*. In addition, this fails to take into account the position of a speaker who is not attributing their evaluation to someone else directly but is instead denying their own ability to make an evaluation due to their lack of knowledge, authority or interest.

Moderator: *Right ok so who do you think it would be aimed at what kind of person do you think.*

Speaker A: *I think I wouldn't say from a male perspective that's something I'd like maybe a female I don't know but er other than that that I wouldn't really have a view I couldn't really say.*

Moderator: *Ok what do you think?*

Speaker A: *I think if it was twenty percent smaller it would definitely appeal to men I can't really say for women.*

Here the speaker gives an evaluation but then follows that up by saying that they are not in a position to provide an informed opinion. One way in which these kinds of utterances have been interpreted previously is from a politeness perspective. As noted in section 2.1.2, this denial of knowledge or authority can be used to lower the strength of an evaluation (Pomerantz 1984). However if the claim of ignorance or lack of knowledge was for politeness reasons, it might be expected that they would occur prior to the evaluation acting to delay the point at which it is given. However, in both of the examples given above, the denial of authority or knowledge comes after the evaluation.

The Attribute branch also seems to miss some of the focus of the following utterance

Speaker A: *I think older ladies might think oh that's really nice that looks really pretty but I wouldn't have my deodorant looking pretty on the bathroom side.*

Here, the evaluation of it looking 'really nice' and 'really pretty' is attributed to 'older ladies'. However the speaker then goes on to say that these are not attributes that are relevant to themselves, or what they are interested in from their deodorant pack. While the initial attribution has been accounted for, the fact that this is not relevant to the speaker has been missed and this is obviously an important point. I would argue, therefore, that there are two distinct types of Relevance. Firstly there is the question of whether or not the speaker sees themselves as being able to make an evaluation based on their knowledge or experience. This would cover evaluations or phrases such as:

- a) *[Evaluation...] although I can't speak for women.*
- b) *[Evaluation...] but younger kids might like it I suppose.*
- c) *I'm the wrong person to ask.*
- d) *I couldn't really say.*

Secondly there is the issue of whether the subject matter is something that the speaker sees as relevant to themselves. If it is not, they may still make an evaluation but then mitigate it by saying that that it is not relevant to them anyway.

- a) *[Evaluation...] but that wouldn't interest me.*
- b) *[Evaluation...] though I don't like anything like that anyway.*
- c) *[Evaluation...] but then that just isn't the sort of thing that matters to me.*

Relevance would sit within the Engagement part of the Appraisal Theory Framework and like Attribute it would be part of the Heteroglossic branch. The two distinct types of Relevance could be named Relevance: Appropriate and Relevance: Importance.

7.4.2 Neutral Evaluations

As noted in section 7.2.2, the concept of neutral evaluations is one that does not appear to be fully catered for within the Appraisal Theory Framework. Although the majority of evaluations are relatively easy to assess in terms of their polarity and position in the framework, there are other utterances which the framework does not account for. While both Martin and White (2005) and Hunston and Thompson (2003) talk of evaluation as being primarily positive or negative, the concept of neutrality is discussed within the literature. Page (2003) talks of neutral evaluations in relation to the way in which they might be affected by their co-text and Coffin and O' Halloran's (2006: 78) paper also talks of the way in which 'seemingly neutral representations of certain groups may come to be interpreted in quite negative ways by a target audience'.

Neutral evaluations can be realised in several different ways. The most obvious is where the subject matter has easily recognizable positive and negative extremes, within which a neutral evaluation can be located. One of the subject matters under discussion was whether or not the stimulus looked as if it would be for an expensive or a cheap product. In most cases evaluations that indicated that it was a premium pack or that it looked expensive or top of the range were tagged as [ATTITUDE: APPRECIATION: REACTION: QUALITY: POSITIVE]. Where it was evaluated as looking cheap or Tesco's own or bottom of the range it was tagged as [ATTITUDE: APPRECIATION: REACTION: QUALITY: NEGATIVE]. There were some cases where it was clear from the surrounding evaluations that the participant saw the perceived expense as being unnecessary and therefore negative, and that being cheaper was actually a positive evaluation. However, the idea of different reading positions has been discussed above in section 5.4 and these evaluations were tagged accordingly. For most evaluations on this subject matter it was unproblematic to tag the evaluation as either positive:

- a) *Oh you'd expect it to be a bit more.*
- b) *I would probably say that it would be more expensive.*
- c) *I think more expensive because of the shape.*

Or negative:

- a) *Makes it look cheap.*
- b) *It's like a value pack.*
- c) *Value Tesco Value.*

However there were also some evaluations on this topic where it seemed that the speaker was trying to make their evaluation as neutral as possible with regards to the perceived expense of the brand that the stimulus has been designed for:

- a) *I'd probably say about middle.*
- b) *Not expensive but not mega cheap either.*
- c) *No I think that would be average priced.*

These instances would seem to be, semantically at least, examples of a more neutral evaluation. In some cases, prosodic features could be used to make an argument for it being one rather than the other. Painter (2003: 204) argues that 'voice quality, intonation and facial expression evoke appreciation in an otherwise neutral utterance' but this would not always be helpful and goes beyond the scope of Appraisal Theory. This situation was also replicated within areas of discussion such as the masculinity or femininity of the stimulus and whether or not the pack was aimed at younger or older people. In these instances it would seem to make sense to add a third strand to the Attitude: Polarity branch of the framework. So rather than having to choose between Attitude: Positive and Attitude: Negative:



Figure 27 The Attitude: Polarity branch of the Appraisal Framework (Martin and White 2005).

It might make more sense to enable the researcher to tag an evaluation for Affect, Judgement or Appreciation but to indicate that either the speaker has chosen to keep their

position neutral, or that the analyst has been unable to ascertain the direction of their evaluation.

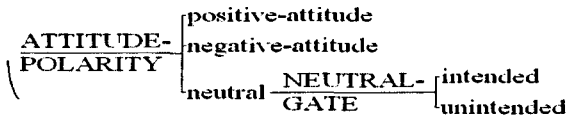


Figure 28 Amended Attitude: Polarity branch of the Appraisal Framework.

In addition to subject matters with obvious extremes, there were also instances of evaluations which appeared neutral within topics of discussion which were less obviously polarised as being positive or negative:

- a) *That that would look ok.*
- b) *I might buy it I I mean I might not it depends on erm.*
- c) *Yea it's not good not bad.*

Here it is not immediately obvious to see how these could be analysed as either positive or negative evaluations. Intonation could again be analysed but this would not necessarily be conclusive.

When the polarity of an evaluation is dependant on the local linguistic context, but that context isn't provided, this could also be seen as a neutral evaluation.

Moderator: *Would you see there as being different parts or would you just see it as being a single thing?*

Speaker A: *I would see it see it as being the whole thing.*

Without further comments suggesting that this is positive or negative, it is impossible to know whether the speaker sees the property of 'being a whole thing' as positive or negative. Hood (2004) also comes up against this problem with the use of the word 'traditionally'. On one occasion she tags it as negative attitude in 'a context in which the writer takes a strongly positive stance towards a progressive pedagogy (Hood 2004: 113).

Subsequently, however, there is then another use of the term 'traditionally' which is 'not supported by other positive (or negative) references in the co-text'. She goes on to argue that 'because of this lack of additional clues to an attitudinal interpretation it is not coded as attitudinal in this data' (Hood 2004: 113). So although it might be possible to look to both the local and larger discourse context, to see if there was any indication as to its intended polarity, this context might not be provided. If it is provided, but only at the larger level, it might not help determine the polarity of an individual utterance.

Another type of neutral comment is when a speaker makes an evaluation that has a negative polarity towards a negative evaluation.

- a) *It's not bad.*
- b) *I don't dislike it.*

This cannot be taken to mean that the speaker thinks that the stimulus is good but neither can it be taken to mean that they think it is bad. This could be interpreted in one of two ways:

- It's not bad (but it's not far from being bad).

Or

- It's not bad (as a fixed expression meaning that it's actually quite good).

Again, some clues as to the intended meaning of such evaluations could be found by looking at their intonation but this goes beyond the scope of Appraisal Theory. This discussion chapter set out to explore many of the issues and questions that presented themselves throughout the thesis. Having attempted to do that, Chapter 8 will provide a brief conclusion, summarising each of the preceding chapters before suggesting both the study's strengths and weaknesses and potential areas of further study.

8. Conclusion

This thesis set out to assess different ways in which a linguistic approach might be used to understand evaluative language within the context of market research interviews and new product design. In Chapter 1, I set the context for this research by introducing and evaluating research in areas relating to consumer behaviour, new product development and consumer insight research. The reasons for consumers wanting or needing products were explored, followed by an introduction to different theoretical attempts to map out this decision making process. The way in which products may or may not be seen in relation to different attributes was then introduced. This was followed by a discussion of the strength of the relationship between reported and actual behaviour. This initial chapter concluded with a brief summary of the literature regarding the use of different qualitative methodologies in the New Product Development field and a review of the way in which some kinds of linguistic analysis have started to be used within a market research setting. Even at this initial stage, it became apparent that while developments in fields such as linguistics were slowly seeping through into the collective conscience of commercially practicing market researchers, there was plenty of scope for developing a sound, linguistics based, methodology that might be able to provide different insights into consumers' thoughts and behaviours.

However, a linguistic approach to understanding and measuring evaluation is far from a new endeavour and therefore Chapter 2 set out to provide the linguistic backdrop against which this research is positioned. This Chapter introduced approaches from four core areas of linguistics encompassing, Discourse, Grammar, Lexis and Phonology. Although I acknowledged that a phonological analysis might provide more detail to assist the analysis of evaluative force, the relationship between prosody and affect is still unclear. The extra time that would have been required to produce transcripts at this level of detail meant that it was decided at this stage that a prosodic analysis would not form part of the research methodology. The discussion of the literature within Discourse Analysis was the main trigger for looking to account for the evaluative force that might be inherent in an utterance due to its place in the turn taking structure. The need to account for this was

further strengthened after the introduction and discussion of concepts from Politeness Theory. Finally, the chapter closed with a brief review of the literature from the field of non-verbal communication. While it was clear that there were some occasions where an analysis of body language or facial features might be beneficial in measuring the evaluative force of an utterance, it also quickly became apparent that this would entail enough work for a PhD in its own right and was therefore seen as being beyond the scope of this research project. The exploration of previous linguistic research in this field led into Chapter 3 and the introduction and discussion of two frameworks of analysis that have been developed within Applied Linguistics. Systemic Functional Grammar is a significant approach in this field and was introduced and discussed in relation to the newer and less well developed framework of Appraisal Theory. While Systemic Functional Grammar would be useful in analysing certain aspects of language use, it was obvious that many of the branches and categories of the Appraisal Theory Framework mapped directly onto the kinds of areas that would be of interest to market researchers, and specifically those working in a new product development field. Having provided the context for the research and discussed potential analytical approaches, Chapter 4 explored many of the different ways in which the data collection process can be carried out. Different elicitation methods were introduced and discussed and issues pertaining to transcription methods were also discussed here. A small scale pilot study was performed to test both the data elicitation and data analysis techniques and by the end of this chapter the main methodologies for this research had been determined. At this point it was also decided that Appraisal Theory would be the most appropriate analytical framework to help analyse the strength of different evaluations.

Chapter 5 then detailed the exact methodology that was used to elicit the data for this research project. The different stages that were used to develop the prototypes were introduced before a detailed examination and justification of the criteria for tagging each specific part of the framework was presented. This chapter also presented the additional layer of analysis that was developed using approaches from Conversation Analysis and Politeness Theory to try and capture the inherent force contained within evaluations due to their role in the turn taking structure. These layers combined to build up the overall

framework for measuring the evaluative force of an utterance. The CorpusTool software that was used to tag and query the data was also introduced in this chapter before it concluded with an inter-rater reliability analysis.

Once the data had been collected, transcribed and tagged, they were then displayed in relation to several key questions that were raised in Chapter 6. The data were described in relation to preference, place in the turn taking structure, content and non content turns and the different weakening and strengthening strategies that the evaluations contained. Presenting the data in this way facilitated a preliminary discussion towards answering the questions put forward in the introduction to the chapter.

These key questions were then re-examined in more detail in Chapter 7 which was comprised of the main discussion sections. It was established that while a surface level analysis might have produced evidence of positive evaluations for the least liked pack, a closer linguistic analysis showed that each of these examples also contained strategies and approaches that acted to minimise the force of these apparent positive evaluations. Similarly it was found that negative evaluations for this stimulus were nearly twice as likely to have been strengthened, than positive evaluations. Perhaps surprisingly, it was shown that, on the surface, overall preference levels towards the pack had little impact on the percentage of evaluations that were either weakened or strengthened. Approximately half of all evaluations were weakened in comparison to a sixth of evaluations being strengthened. However, when a more detailed analysis was carried out, taking into account the polarity of the evaluation, its role in the turn taking structure and also the different categories of the Appraisal Theory Framework, it was possible to spot more interesting trends and patterns. Understanding these patterns made it possible to infer why one pack proved more favourable than the other. It was possible to surmise that the stimulus Mouse had a very strong impact at the first moment of truth and that this compensated for any problems with the actual design or dimensions of the pack. Similarly it was apparent that for Egypt, the problems with the physical dimensions and usability of the pack proved too strong a hindrance, despite some positive evaluations regarding its design and aesthetic appeal. Insights such as these, that might not have been

possible without a detailed linguistic analysis, suggest that there might be significant advantages in using an approach that can systematically measure the evaluative force of any utterance.

Chapter 7 then further discussed ways in which this methodology and approach could be applied to the consumer research industry. Two areas where I felt that the Appraisal Theory Framework might be extended were then also introduced. These extensions aimed to account for evaluative markers relating to the relevance of the evaluation to the speaker and also to try and accommodate evaluations which were neutral, rather than positive or negative, in nature. The following section discusses ways in which the analytical approach detailed in this thesis can be applied in a real world commercial setting. Finally, it concludes by exploring some of the strengths and limitations of this study and then looking at potential areas for further research.

8.1 Applying linguistic approaches to commercial practice

This section attempts to answer to research question 3 posed in Chapter 1:

How could manufacturers of products use knowledge of these considerations when designing products and packaging?

In chapter 1 I discussed the ways in which consumer insight can be integrated into the NPD process, particularly with regards to the use of focus groups and interviews for both identifying consumer needs and also testing and refining prototypes. I argue that it is at these stages that knowledge of the social and interactional influences on evaluation has the potential to attain different kinds of insight than are currently achieved. As discussed in section 2.1.1, and drawing on personal experience having worked in the industry for several years, there is often little time within a commercial setting for a formal content analysis of the data. Indeed, it is quite common for transcripts of the groups or interviews not to be made at all; the researcher may simply listen back to the recordings or make use of their own, or another researcher's notes. I would argue that in comparison, an

approach that not only systematically explores all of the evaluative content, but also takes into account the social and structural elements of communication has the potential to offer more. As can be seen in the discussion in chapter 7, a surface level analysis of the interviews that were carried out for the stimulus Mouse and Egypt might have produced very different results from the level of detail that was achieved by using the methodology proposed in this research. Knowledge of the social and interactional elements of communication enables the researcher to better interrogate their data and therefore gain a different level of insight.

However, two barriers to the use of this approach within a commercial setting are those of time and money. The time between carrying out the fieldwork interviews and presenting the final debrief to the client can be as little as two weeks and in these circumstances there simply would not be time to carry out the transcription, tagging and analysis that would be required.

The nature of the commercial research would also have an impact on how much of an advantage a full linguistic analysis would provide. Within the consumer insight industry there is a wide range of topics and client needs that are explored. These range from projects very similar to this research, exploring FMCG product packaging design, to much higher level strategic projects. For instance, this could be exploring attitudes to government policies or towards concepts such as what does 'environmentally friendly' or 'natural' mean when associated with products or services. For those projects exploring tangible 'things' where it is important to understand the small nuanced differences in attitudes towards broadly similar stimulus, a detailed analytical approach would be more useful. For those looking to explore broad consumer attitudes or trends or responses to communications material, it is perhaps less relevant.

One way round the time and money issue would be to automate as much of the process as possible. Developments in voice recognition software or even the use of stenographers means that it would be possible to have a full transcript of the group or depth interviews carried out in near real time. It is likely that this would need to be tidied up or corrected

to some extent but it would still be suitably fast. As discussed in section 5.5, a database of evaluations was built as part of this research study. This database contained each of the different evaluative units that occurred and detailed which of the categories of the Appraisal Theory framework they had been assigned to. It is possible to write software that would automatically connect a database such as this to the corpus tool software. It would then be possible to feed raw transcripts into the software and almost instantaneously look at the way in which evaluative language had been used. If the transcripts were tagged for the different speakers, the turn-taking layer element could also be automated. As discussed throughout section 5.4 there are likely to be incidences where the polarity or meaning of the evaluation is not immediately obvious and therefore there is a limit to the extent to which the whole process could be automated. However, as the database grew it is likely that the level of automation would increase. While I wouldn't argue that this kind of automated output could ever replace the analysis that currently takes places within the consumer insight industry, it could still prove a useful addition.

It would, however, be doing consumer insight professionals a disservice to suggest that they are completely unaware of the social and structural elements of evaluation. Many of the areas that this thesis has covered are things that they pick up on intuitively. It is not uncommon for a moderator to make an observation such as *'they said they like it but I'm sure they don't really'* or *'they're just being polite, you can tell they're not interested in it'*. In instances such as this they are subconsciously picking up on the kind of weakening strategies that have been discussed throughout this thesis. Good, experienced moderators or observers gain more insights from their interviews for the very reason that they are well attuned to listening out for these elements, even if they are not aware of the theories that lie behind them. Even where this is the case though, it is simply not possible for an individual to take account of all of the different elements from each of up to eight respondents and as such I argue that the approach proposed in this research would always have some benefit.

In addition to the main way in which this approach could be used it is also proving useful in three other ways. Throughout the thesis it has been argued that this approach adds to the analytical element of the research process. From a personal experience I would argue that a knowledge of the social and structural elements of communication can also be used to improve the moderation of research interviews and groups. Being aware of the different weakening and strengthening strategies that respondents can draw on makes it possible to challenge these in situ. For instance the following exchange is taken from a recent interview carried out as part of my professional role as research consultant.

- Respondent 6: *Mmm so they're thinking of doing this bringing out I'm quite excited I'd definitely try it.*
- Respondent 4: *Yea I know what you mean I do quite like the idea*
- Moderator: *Ok [respondent 4] you said you quite like it, is that just because [respondent 6] loves it or do you really like it?*

Having observed the weakening strategies employed by Respondent 4, I then challenged them to check the extent to which their response was being guided by the structural and social context. So I would argue that a moderator who is explicitly aware of these issues can not only use that knowledge in the analysis stage, but also whilst carrying out the fieldwork interviews.

The second way in which knowledge of the social and structural elements of communication can be applied to industry is through the training of new or inexperienced moderators. Due to the diverse skill set that is required to work in the industry, there is not a single entry level qualification that new qualitative researchers are required to have. This means that many new starters have no formal training in qualitative moderation. The professional body, the Association of Qualitative Researchers (AQR), provides training courses but many moderators simply learn on the job by watching more experienced moderators. In addition the training courses run by the AQR are lacking in a formal linguistic focus.

The final way in which this knowledge has been applied in a commercial setting is in instances where the commercial client has been viewing the groups. When the client wishes to observe the groups first hand, viewing facilities are used whereby the client and any other researchers sit behind a two way mirror to watch the proceedings in real time. Where the focus of the research is to determine which packaging solution, new product development or piece of communication is best liked by consumers, the client will often have a preference before the research has started. They may then cherry pick comments from consumers which appear to indicate a preference for this favourite. In many instances the moderator may feel that despite the apparent positive comments, that particular stimulus is not the preferred item. They may be picking up on weakening strategies that a respondent has used and knowledge of these elements, as set out in this research, would enable the moderator to explain this fact to the client and justify why they might recommend a different stimulus.

For example if clients had been observing the research groups carried out for this study, and had a preference for Egypt prior to the interviews taking place, they would have been able to find many examples of apparent positive comments and evaluations. However, due to my knowledge of the weakening strategies that were being used, it would have been possible to explain to the client that Mouse may actually be the preferred option. While I would not suggest that it is possible to keep track of all of the different weakening and strengthening elements going on within an interview, I do argue that it is possible to observe them in individual evaluations that a client might present as evidence in favour of their pre-existing view point.

8.2 Strengths and Limitation of the study

This study set out to develop a deeper understanding of the way in which evaluative language is formed, with a view to developing a framework for analysing evaluative language that is elicited in market research interviews. In seeking to answer my research questions, this research has combined theories and approaches from Politeness Theory and Conversation Analysis, in conjunction with the developing field of

Appraisal Theory. I have set out a framework for analysis that allows the researcher to systematically account for each of the different social and interactional constructs that have an impact on the strength and polarity of an evaluation.

8.2.1 Contributions to the Knowledge

Appraisal Theory has been continually developed since it was first proposed, and as a result of this research I have suggested two extensions to the framework. Firstly, I suggest the need for an additional sub branch within the ENGAGEMENT strand of the framework, to account for how relevant the subject matter is to the speaker. Secondly I argue that the polarity of evaluations is not always clear, by design or otherwise, and therefore the analyst should be able to tag the evaluations appropriately. In addition to taking an existing framework and developing it for a new purpose, this thesis also contributes to the wider understanding of 'evaluation', through the development of a Scale of Importance for individual turns, with regards to the 'weight' that should be assigned to them due to their place in the turn taking structure. This study has taken an existing analytical approach and used it for a significantly different purpose, with real world, commercial applications. During my placement there was considerable interest in the analytical framework that I was developing. I am now continuing to develop and use this framework, for analysing consumer evaluations, in a commercial setting for a leading insight and innovation consultancy.

8.2.2 Limitations

Although I argue that this research has carried out new and interesting work, there are inevitably limitations. Firstly, this research is the work of a single analyst and as such, any flaws or inaccuracies in the transcription, tagging or analysis of the data, remain largely unchecked. Secondly, despite collecting over thirty hour's worth of data, only a smaller sample of this was transcribed, tagged and analysed in detail. Due to the time constraints of this project, it was felt that it would not be possible to carry out a full analysis of all of the data. As detailed in Chapter 6 only data from the most and liked stimulus was used. In addition only half of the interviews were transcribed and analysed

in full, primarily for time and practicality reasons but also due to their suitability for use. I.e. on four occasions only one of the respondents arrived and though the interview was still carried out as an individual depth interview, the different research methodology meant the data could not be incorporated into the study. By using the stimuli that were the most and least liked, it is hoped that this has ensured that a full spread of both positive and negative evaluative language has been analysed in this study. As discussed in section 5.5, an obvious potential weakness for a study of this kind is that it requires some subjective decision making with regard to the tagging of the data. This potentially weakens any subsequent claims or findings. To counteract this, an inter-rater reliability exercise was carried out, though again it could be argued that more co-raters or more data could have made this test more robust.

It could perhaps be argued that no full account of the evaluative strength of an utterance can take place, without at least some consideration of the prosodic and non verbal aspects of communication. However, while these areas were briefly introduced and discussed in sections 2.4, a framework combining all of these elements would go beyond the scope or possibilities of one PhD thesis. Therefore, although it would have been interesting to try to account for these areas, it simply was not possible within the boundaries of this research.

8.3 Areas for Further Research

As noted in section 4.1.2.1, it is possible that some physical, concrete properties have an intrinsic link to more abstract properties. By acting as a trigger to more abstract properties or attributes there might be a causal relationship between them. A corpus study could be used to analyse the collocation frequency of certain words or alternatively participants could be presented with a list of twenty concrete and twenty abstract properties and then asked to match them up in pairs. By triangulating the research and using both methodologies it may be possible to conclude that there is a link between concrete and abstract attributes. This would then perhaps challenge Beck-Larsen and Nielsen's (1999) claim that concrete attributes are less important. If they act as triggers

or precursors to more abstract attributes then they could be viewed as being of equal or even, higher importance. Other potential areas of interest that were felt to be beyond the scope of this study included an analysis of prosodic features such as tone pitch and pace and also the inclusion of non verbal features such as body language and facial expression. With sufficient expertise, resources and time, it might eventually be possible to develop a framework that would encompass an analysis of each of these different communication channels to provide an even greater understanding of the evaluative force of any given utterance. However, while this research only concentrated on one of these channels, it is hoped that it has contributed to the understanding of evaluative language and provided a framework that can help to measure the evaluative force contained within any given utterance.

References

Agar, M. (1996) *The Professional Stranger: An Informal Introduction to Ethnography*. Boston: Houghton

Ajzen, I. (1988) *Attitudes, Personality, and Behaviour*, Milton Keynes: Open University Press.

Ajzen, I. (1991) The Theory of Planned Behaviour. *Organizational Behaviour and Human Decision Processes*. 50 (2), pp. 179-211.

Ajzen, I. (2002) Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, 32 (4), pp. 665-683

Ajzen, I. and Fishbein, M. (2005). The Influence of Attitudes on Behaviour. In: Albarracín, D., Johnson, B., Zanna, M. (eds.) *The Hand Book of Attitudes*. London: Routledge. pp. 173-221.

Albarracín, D., Johnson, B. and Zanna, M. (2005) *The Hand Book of Attitudes*. London: Routledge.

Algeo, J. (1990) *It's a Myth, Innit?* Politeness and the English Tag Question. In: Ricks, C. and Michaels, L. (eds.) *The State of the Language*. Berkeley: University of California Press. pp. 443-50.

Antonides, G. (1991) *Psychology in Economics and Business*. Dordrecht: Kluwer Academic Publishers.

Areni, C. S. and Kim, D. (1993) The Influence of Background Music on Shopping Behaviour: Classical versus Top-Forty Music in a Wine Store, *Advances in Consumer Research*, 20, pp. 336-340.

Areni, C. S. and Kim, D. (1994) The influence of In-store Lighting on Consumers' Examination of Merchandise in a Wine Store. *International Journal of Marketing Research*, 11, pp. 117–125.

Armitage, C. J. and Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, 40, pp. 471–499.

Assael, H. (2004) *Consumer Behaviour - A Strategic Approach*. Boston: Houghton Mifflin Company.

Atkinson, J. and Heritage, J. (eds.) (1984) *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.

Baker, M., Francis, G. and Tognini-Bognelli, E. (1993) *Text and technology: In honour of John Sinclair*. Amsterdam: J. Benjamins.

Bakhtin, M. M. (1935 [1981]). *The Dialogic Imagination: Four Essays*. Edited by M. Holquist, translated by C. Emerson and M. Holquist. Austin: University of Texas Press.

Bakhtin, M. M. (1953[1986]). The problem of speech genres. Translated by Vern W. McGee. In *Speech Genres and Other Late Essays*, C. Emerson and M. Holquist (eds.), 60–102. Austin, TX: University of Texas Press.

Bandura, A. (1997) *Self-efficacy: The exercise of control*. New York: Freeman.

Barrata, A. M. (2009) Revealing stance through passive voice. *Journal of Pragmatics*. 41 pp. 1406 -1421.

Barton, E.L. (1993). Evidentials, argumentation, and epistemological stance. *College English*, 55 (7) pp. 745 –769.

Beck-Larsen, T. and Nielsen, N. A. (1999) A comparison of five elicitation techniques for elicitation of attributes of low involvement products. *Journal of Economic Psychology*, 20, pp. 315-41.

Beebe, L. and Takahashi T. (1989) *Do You Have a Bag? Status and Patterned Variation in Second Language Acquisition*, In: S. Gass, et al (eds) *Variation in Second Language Acquisition: Discourse and Pragmatics*, Clevedon: Multilingual Matters, pp. 103-125.

Berman, R. (2005) Introduction: developing discourse stance in different text types and languages. *Journal of Pragmatics*, 37 (2), pp. 105-124.

Berry, L.L, Carbone, L.P. and Haeckel, S.H. (2002) Managing the total customer experience, *MIT Sloan Management Review*, Spring, pp. 85-89

Bertrand, K. (2002) Wake up Your Product Category with 'Shapely' Packaging. In: Underwood, R.L. and Klein, N.M. (eds) *Packaging as brand communication: effects of product pictures on consumer responses to the package and brand*, *Journal of Marketing Theory and Practice*, 10 (4), pp. 58-68.

Biber, D. and S. Conrad. (2002). 'Register variation: A corpus approach'. In D. Schiffrin, D. Tannen and H. Hamilton (eds) *The handbook of discourse analysis*. M.A.: Blackwell. pp. 175-196.

Biber, D. and Finegan, E. (1989) 'Styles of stance in English: lexical and grammatical marking of evidentiality and affect'. *Text*, 9 (1), pp. 93-124.

Birdwhistell, R.L. (1952) *Introduction to Kinesics*. University of Louisville Press

Birdwhistell, R.L. (1970) *Kinesics and Context Volume 1*. London: ALP

Birdwhistell, R.L. (1971) *Kinesics and Context Volume 2*. London: ALP

Blum-Kulka, S., Danet, B., and Gherson, R. (1985). The language of requesting in Israeli society. In: F. Forgas (ed.) *Language and social situation*. New York: Springer-Verlag. pp. 113–136.

Bogue, J. and Ritson, C. (2006) Integrating consumer information with the new product development process: the development of lighter dairy products. *International Journal of Consumer Studies*, 30, pp. 44-54.

Bosmans, A. and Baumgartner, A. (2005) Goal Relevant Emotional Information: When Extraneous Affect Leads to Persuasion and When It Does Not. *Journal of Consumer Research*. 32 (3), pp. 424–434.

Breal, M. (1897) *Essai de sémantique*, éd Gérard Monfort, Brionne In Louw, B (1993) *Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies* In Baker, M., Francis, G. and E. Tognini-Bonelli (Eds), *Text and Technology: In Honour of John Sinclair*, Amsterdam, John Benjamins, pp. 157-176.

Brown, P. and Levinson, S. (1987) *Politeness: Some Universals in Language Use*. Cambridge: Cambridge University Press.

Bryman, A. (2001) *Social Research Methods* Oxford: Oxford University Press

Buchanan, T. et al. (2000) Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Cognitive Brain Research*, 9 (3), pp. 227-238.

Bybee, J. L. and Fleischman, S. (1995) *Modality in Grammar and Discourse*. Amsterdam: J. Benjamins.

Cameron, D. (2001) *Working with Spoken Discourse*. London: Sage.

Cameron, D. (2007) *The Myth of Mars and Venus: Do Men and Women Really Speak Different Languages?* Oxford: Oxford University Press.

Cameron, D., McAlinden, F. and O'Leary, K. (1989) Lakoff in Context: The Social and Linguistic Functions of Tag Questions. In: Coates, J. and Cameron, D. (eds.) *Women in Their Speech Communities: New Perspectives on Language and Sex*, London: Longman. pp. 74-93.

Carletta, J. (1996). Assessing agreement on classification tasks: The Kappa statistic. *Computational Linguistics*, 22 (2), pp. 249-254.

Carlson, R. (1991). Synthesis: modelling variability and constraints. In: *Proceedings of the Eurospeech conference*, Genova, Italy, 1991, pp. 1043-1048.

Carlson, J (1989) *Moments of Truth*. London: Collins.

Carspecken, P. (1996) *Critical Ethnography in Educational Research: A Theoretical and Practical Guide*. London: Routledge.

Chafe, W. L. (1982) Integration and involvement in speaking, writing, and oral literature. In: D. Tannen (ed.) *Spoken and Written Language: Exploring Orality and Literacy*. Norwood: Ablex. pp. 35-54.

Chafe, W. L. (1986) Evidentiality in English Conversation and Academic Writing. In: W.L. Chafe and J. Nichols (eds), *Evidentiality: The Linguistic Coding of Epistemology*. Norwood: Ablex, pp261-272.

Chafe, W. L. and Nichols, J. (1986) *Evidentiality: The Linguistic Coding of Epistemology* London: Ablex Publishing Corporation.

Channell, J. (1994) *Vague Language*. Oxford: Oxford University Press.

Channell, J. (2000) 'Corpus-based analysis of evaluative lexis' In: Hunston, S. and Thompson, G. (eds.) (2000) *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press. pp. 38-54.

Charmaz, K. (2000) Grounded Theory: Objectivist and Constructivist Methods. In: Denzin, N. and Lincoln, Y. (eds.) *Handbook of Qualitative Research*. London: Sage. pp. 509-536.

Choy, K.L. (2007) Design of a generic product and process development system – a hybrid case-based reasoning approach. *International Journal of Product Development*, 4, pp.310-329.

Coates, J. (1996) *Women Talk*. Cambridge, MA: Blackwell.

Coffin, C., and O'Halloran, K. (2006) The roles of appraisal and corpora in detecting covert evaluation. *Functions of Language*, 13 (1), pp. 77-110.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, pp. 37–46.

Conrad, S. and D. Biber. (2000) 'Adverbial marking of stance in speech and writing'. In: Hunston, S. and Thompson, G. (eds.) (2003) *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press. pp. 56-73.

Cook, G. (2001) *The Discourse of Advertising*. London: Routledge.

Cooke, R. and Sheeran, P. (2004) Moderation of cognition-intention and cognition-behaviour relations: A Meta-analysis of properties of variables from the theory of planned behaviour. *British Journal of Social Psychology*, 43, pp. 159-86.

Corey, S.M. (1937) Professed Attitude and Actual Behaviour. *Journal of Educational Psychology*, 28, 271-280. In: Summers, G. (ed.) (1970) *Attitude Measurement* Chicago: Rand McNally.

Cortazzi, M. (1991) *Primary teaching, how it is: A narrative account*. London: David Fulton

Cortazzi, M. and Jin, Z. (2003) Evaluating Evaluation in Narrative. In: Hunston, S. and G. Thompson (eds.) *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press pp. 121-141.

Coulthard, M. (1992) *Advances in written text analysis*. London: Routledge.

Croft, W. (2003) *Typology and Universals*. Cambridge: Cambridge University Press.

Dascal, M. (2001) 'Reputation and refutation: negotiating merit'. In: E. Weigand and M. Dascal (eds.) *Negotiation and Power in Dialogic Interaction*. Amsterdam: John Benjamins. pp. 3-17.

Davies, A., and Cline, T., W. (2005) A consumer behaviour approach to modelling monopolistic competition. *Journal of Economic Psychology*, 20 (6), pp.62-75.

Davis, J. (1976) *Elementary Survey Analysis*. Englewood Cliffs, NJ: Prentice Hall.

Dendale, P. and Tasmowski, L. (2001) Introduction: evidentiality and related notions. *Journal of Pragmatics*, 33 (3) pp. 339-348.

Denzin, N. and Lincoln. Y. (eds.) (1994) *Handbook of Qualitative Research* Thousand Oaks CA: Sage.

De Vos, A. (1998) *Research at Grassroots*. Pretoria: J.L. Van Schaik.

Doyle, M. (2002) Consumers Want 'Packaging to Go'. *Journal of Consumer Behaviour*, 2 (3), pp. 220-232.

Duncan, S. and Fiske, D. (1977) *Face-to-face Interaction: Research, Methods, and Theory*. New York: Erlbaum Associates.

Eggin, S. and Slade, D. (2001) *Analysing Casual Conversation*. London: Continuum.

Engel, J. F., Blackwell R. D. and Miniard, P.W. (1994) *Consumer Behaviour*. Fort Worth, Tx: The Dryden Press, Harcourt Brace College Publishers.

Fairclough, N. (1989) *Language and Power*. London: Longman.

Fairclough, N. (1993) *Discourse and Social Change*. Cambridge: Polity Press.

Fairclough, N. (1995) A Reply to Henry Widdowson's 'Discourse analysis: a critical review' *Language and Literature*, 3 (9), pp.49-56.

Fairclough, N (2003) *Analysing Discourse*. London: Routledge.

Farooq, M. (1999) 'Analysis of a Casual Conversation Based on Francis and Hunston's Model'[Online]. Available at www.bham.ac.uk/CELS/CELS%20pages/essays/farooq4.pdf

Fazio, R., H., Williams, C., J. and Powell, M.C. (2000) Measuring Associative Strength: Category-Item Associations and Their Activation from Memory, *Political Psychology*, 21 (1), pp. 7-25.

Fazio, R. H. (1986) How Do Attitudes Guide Behaviour. In: R.M. Sorrentino and E.T. Higgins, (eds.) *The Handbook of Motivation and Cognition*. Wiley: Chichester. pp. 142-168.

Fazio, R.H. and Zanna, M. P. (1981) *Direct Experience and Attitude-Behaviour Consistency*. New York: Academic Press.

Field, J. (2003) *Psycholinguistics: A Resource Book for Students*. London: Routledge.

Fielding, N. and Thomas, H. (2001) *Qualitative Interviewing*. In: Gilbert, N (ed.) *Researching Social Life*. London: Sage. pp.123-145.

Firth, J.R. (1935) The Technique of Semantics. Transactions of the Philological Society. In Cortazzi, M. and Jin, Z. (2003) Evaluating Evaluation in Narrative. In: Hunston, S. and Thompson, G. (eds.) *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press. pp 1.21-141

Fishbein, M. and Ajzen, I. (1980) *Understanding Attitudes and Predicting Social Behaviour*. Englewood Cliffs, NJ: Prentice Hall.

Fitch, K.L. and Sanders, R.E. (2005) *Handbook of Language and Social Interaction*. New York: Erlbaum.

Flick, U. (2002) Qualitative Research- State of the Art. *Social Science Information*, 41, pp. 5-24.

Flick, U. (2009) *An Introduction to Qualitative Research*. London: Sage.

Fowler, R. et al. (1979) *Language and Control*. London: Routledge.

Foxall, G. and Johnston, B. (1987) Strategies of User-initiated Product Innovation *Technovation*, 6 (1), pp. 77-102.

Francis, G. and Hunston, S. (1992) *Analysing Everyday Conversation*. In: Malcolm Coulthard (ed) *Advances in written text analysis*. London: Routledge. pp. 123-162.

Frankfort-Nachmias, C. and Nachmias, D. (2000) *Research Methods in the Social Sciences*. New York: Worth Publishers.

Garg, N., Inman, J. J. and Mittal, V. (2005) Incidental and Task Related Affect: A Re-Inquiry and Extension of the Influence of Affect on Choice. *Journal of Consumer Research*. 32 (1), pp. 154–159.

Gardner, R. (2001) *When Listeners Talk: Response Tokens and Listener Stance*. Amsterdam: John Benjamins Publishing Company.

Gerhardt, J. and Beyerle, S. (1997) What if Socrates had been a woman? The therapist's use of acknowledgement tokens (mm-hm, yeah, sure, right) as non-reflective means of intersubjective involvement, *Contemporary Psychoanalysis*, 33, pp. 367–410.

Gilbert, N (ed.) (2001) *Researching Social Life*. London: Sage.

Goffman, E. (1967) *Interaction Ritual: Essays on Face-to-Face Behaviour*. New York: Anchor Books.

Gorn, G., Pham, M. T. and Sin, L.Y. (2001). When arousal influences ad evaluation and valence does not (and vice versa). *Journal of Consumer Psychology*, 11, pp. 43–55.

Goodluck, H. (1991) *Language Acquisition: A Linguistic Introduction*. Oxford: Blackwell.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Goodwin, C. (1986). Between and within: Alternative treatments of continuers and assessments. *Human Studies*, 9, pp. 205–17.

Goodwin, C. and Goodwin, M. (1992). Context, activity and participation. In: Auer, P. and di Luzio, A. (eds.) *The contextualization of language*, Amsterdam: Benjamins. pp. 77–99.

Goodwin, C. and Heritage, J. (1990) Conversation Analysis. *Annual Reviews of Anthropology*, 19, pp. 283-307.

Grice, H., P. (1975) *Logic and Conversation* In: Cole, P. and Morgan, J. (eds.) *Syntax and semantics*, New York: Academic Press. pp. 155-190.

Gunter, R. (1974) *Sentences in Dialogue*. Columbia: Hornbeam Press.

Gustafsson, A. and Johnson, M.D. (2003) *Competing in a Service Economy: How to Create a Competitive Advantage through Service Development and Innovation*, San Francisco, CA: Carter.

Hague, P. N, Hague, N. and Morgan, C. A. (2004) *Market Research in Practice: A guide to the Basics*. London: Sterling.

Halfpenny, P. (1979) The analysis of qualitative data. *Sociological Review*, 27 (4), pp. 799 -825.

Halliday, M.A.K (1992) Language as system and language as instance: The corpus as theoretical construct. In: J. Svartvik (ed.) *Directions in Corpus Linguistics: Proceedings of Nobel Symposium*. Stockholm, August 1991. De Gruyter: Berlin. pp. 12-18.

Halliday, M.A.K. (1993) *Language in a changing world*. In M.A.K. Halliday, *Language in a Changing World*. (Occasional Papers, 13.) Canberra: Applied Linguistics Association of Australia.

Halliday, M.A.K. (1994) *An Introduction to Functional Grammar*. London: Edward Arnold.

Halliday, M.A.K and Matthiessen, C.M.I.M. (1999) *Construing Experience through Meaning: A Language-based Approach to Cognition*. London: Cassell.

Hardman, M. J. (1986) Data Source Marking in Jaqi Lanugage'. In: W.L. Chafe and J. Nichols (eds) *Evidentiality: The Linguistic Coding of Epistemology*. Norwood: Ablex. pp261-272.

Heritage, J. (1989) Current Developments in Conversation Analysis. In: D. Roger and P. Bull (Eds) *Conversation*. Clevedon: Multilingual Matters, pp. 21-47.

Hodge, R. and Kress, G. (1979) *Language as Ideology* London: Routledge.

Hodge, R. and Kress, G. (1988) *Social Semiotics*, Cambridge: Polity Press.

Holmes, J. (1983) The Functions of Tag Questions. *English Language Research Journal*, 3, pp. 40-65.

Holmes, J. (1990) Apologies in New Zealand English. *Language in Society*, 19 (2), pp.155-99.

Holmes, J. (1995) *Women, Men and Politeness*. White Plains, NY: Longman.

Holtgraves, T., and Yang, J.-N. (1990). Interpersonal underpinnings of request strategies: general principles and differences due to culture and gender. *Journal of Personality and Social Psychology*, 62, pp. 246-256.

Hood, S., (2004) *Appraising Research: Taking a stance in academic writing* (Unpublished PhD Thesis available from www.grammatics.com).

Hornik, R. et al. (2001). Evaluation of the national youth anti-drug media campaign: *Second semi-annual report of findings*. Philadelphia: The Annenberg School for Communication.

Howard, J. A. and Sheth, J. (1998) *A Theory of Buyer Behaviour*. London: John Wiley.

Hui, M. K., Dube, L. and Chebat, J. (1997) The Impact of Music on Consumers' Reaction to Waiting for Services. *Journal of Retailing*, 73, pp. 87-104.

Hunston, S. and Thompson, G. (eds.) (2003) *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press.

Hutchby, I. Wooffitt, R. (2004). *Conversation Analysis: principles, practices, and applications*. London: Blackwell.

Hyland, K. (1994) 'Hedging in academic writing and EAP textbooks'. *English for Specific Purposes*, 13 (3), pp. 239-256.

Hyland, K. (1998) *Hedging in scientific research articles*. Amsterdam: John Benjamins

Hyland, K. (2000) *Disciplinary discourses: social interactions in academic writing*. New York: Longman.

Hyland, K. (2002) 'Genre: Language, context and literacy'. *Annual Review of*

Applied Linguistics, Discourse and Dialogue, 22, pp. 113-135.

Janda, L. (2006) *Cognitive Linguistics*, Cambridge: Cambridge University Press.

Jaworski, B.J. and Kohli, A.K. (1993) Market orientation: antecedents and consequences. *Journal of Marketing*, 57, pp. 53–70.

Jefferson, G. (1973). A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica*, 9, pp. 47–96.

Jefferson, G. (1978) *Sequential Aspects of Story Telling in Conversation*. New York: The Free Press.

Jin, L. (1992) *Academic Cultural Expectations and Second Language Usage: Chinese Post-graduate students in the UK – A cultural Synergy Model*. (Unpublished Ph.D Thesis. University of Leicester.)

Johne, A. (1994) Listening to the voice of the market. *International Marketing Review*, 11, pp. 47–59.

Jones, R. and Thornborrow, J. (2004) Floors, talk and the organisation of classroom activities. *Language in Society*, 33 (3), pp. 233-423..

Jordan, P. (2000) *Designing Pleasurable Products*. London: Taylor and Francis.

Kardes, F. R. et al. (1993) Brand retrieval, consideration set composition, consumer choice, and the pioneering advantage. *Journal of Consumer Research*, 20 (6), pp. 62-75.

Kelley, S. and Mirer, T.W. (1974) The Simple Act of Voting. *American Political Science Review*, 68, pp. 572-91.

Kelly, G. A. (1955) *The Psychology of Personal Constructs*. New York: Norton.

Kidwell, B. Hardesty, D. M. Childers, T. L. (2008) Consumer Emotional Intelligence: Conceptualization, Measurement, and the Prediction of Consumer Decision Making. *Journal of Consumer Research*. 35, (1), pp. 154–166.

King, S. and McDonald, C, (1996) *Sampling the Universe*, NTC Publications in association with The Market Research Society.

Kitzinger J. (1994) The methodology of focus groups: the importance of interaction between research participants, *Sociology of Health*, 16 (1), pp. 103-21.

Krauss, R., Morrel-Samuels, P. and Colasante, C. (1991) Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, 61 (5), pp. 743-754.

Koller, V. (2008) 'Not just a colour': pink as a gender and sexuality marker in visual communication. *Visual Communication*, 7 (4), pp. 395-423.

Kotthoff, H. (1993) *Disagreement and concession in disputes: On the context sensitivity of preference structures*. Cambridge: Cambridge University Press.

Kress, G. (1985) Ideological structures in discourse. In: T. A. van Dijk (Ed.), *Handbook of discourse analysis: Vol 4. Discourse analysis in society*. London: Academic Press. pp. 27-42.

Krosnick, J. A, and Petty, R. E. (1995) (Eds.) *Attitude Strength: Antecedents and Consequences*. Hillsdale, NJ: Erlbaum.

Krsner, H., (2000) *Negotiating Authority: The Logogenesis of Dialogue in Common Law Judgments*, unpublished PhD dissertation, Linguistics Department, University of Sydney, Sydney.

- Kvale, S. (1996) *An Introduction to Market Research Interviewing*. London: Sage.
- Labov, W. (1966) *The social stratification of English in New York City*. Washington D.C.: Centre for Applied Linguistics.
- Labov, W. (1972) *Language in the Inner City*. Philadelphia: University of Philadelphia.
- Labov, W. (1984) *Intensity*. In: *Meaning form, and use in context: Linguistic Applications*. D. Schiffrin (ed.) Washington: GUP In Biber, D. and Finegan, E (1989) *Styles of Stance in English: Lexical and Grammatical Marking of Evidentiality and Affect*. *Text*, 9 (1), pp. 93-124.
- Lakoff, (1972) *Hedges: a study in meaning criteria and the logic of fuzzy concepts* *Chicago Linguistic Society Papers* Chicago: Chicago Linguistic Society.
- Lampi, M. (1986) *Linguistic components of strategy in business negotiations*. Helsinki: Helsingin Kauppakorkeakoulun Kuvalaitos.
- Landis, J.R., and Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, pp. 159–174.
- Langacker, R. W. (1987) *Foundation of Cognitive Grammar*. New York: SUP.
- LaPiere, R.T. (1934) *Attitudes vs. Actions*. In: Summers, G. (ed.) *Attitude Measurement* Chicago: Rand McNally. pp. 230-37.
- Lee, Y-A., (2007) Third turn position in teacher talk: Contingency and the work of teaching. *Journal of Pragmatics*, 39, pp.180-206.
- Lee, D. (2002) *Cognitive Linguistics: An Introduction*. Oxford University Press: Oxford.

Leech, G. and Svartvik, J. (1996). *A Communicative Grammar of English*. Longman, London.

Leinonen, L. et al. (1997). Expression of emotional-motivational connotations with a one-word utterance. *Journal of the Acoustic Society of America*, 102, pp.1853–1863.

Lim, T. and Bowers, J.W. (1991). Facework: Solidarity, approbation, and tact. *Human Communication Research*, 17, pp. 415–450.

Litosseliti, L. (2003). *Using focus groups in research*. London: Continuum.

Lofgren, M. (2005) Winning at the first and second moments of truth: an exploratory study. *Managing Service Quality*, 15 (1), pp. 102-115.

Long, M, H., and Sato, C. (1983) Classroom foreigner talk discourse: forms and functions of teachers' questions. In: Seliger, H. and Long, M, (Eds.) *Classroom oriented research in second language acquisition*, Rowley, Massachusetts: Newbury House Publishers. pp. 268-285.

Louw, B (1993) Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In: Baker, M., Francis, G. and E. Tognini-Bonelli (Eds), *Text and Technology: In Honour of John Sinclair*, Amsterdam, John Benjamins, pp. 157-176.

Lynch, T. (1996) *Communication in the Language Classroom*. Oxford: OUP.

Lyons, J. (1977) *Semantics*. Cambridge: Cambridge University Press.

Macbeth, D. (2003) Hugh Mehan's "Learning Lessons" reconsidered: on the differences between the naturalistic and critical analysis of classroom discourse. *American Educational Research Journal*, 40, pp. 239–280.

Macken-Horarik M., (2003) Envoi: Intractable issues in appraisal analysis. *Text*, 23 (2), pp. 313–319.

MacLean, L., Mechthild M., and Alma E. (2004) Improving Accuracy of Transcripts in Qualitative Research. *Qualitative Health Research*. 14 (1), pp. 113-123.

Malouf, R. (1995). "Towards an analysis of multi-party discourse" (online at <http://odur.let.rug.nl/malouf/papers/talk.pdf>).

Markkanen, R. and Schroder, H. (1997) *Hedging and discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic texts*. The Hague: Walter de Gruyter.

Martin, J.R (1992) *English Text: System and Structure*. Amsterdam: Benjamins.

Martin, J.R. (1995) Interpersonal Meaning, Persuasion, and Public Discourse: Packing Semiotic Punch', *Australian Journal of Linguistics*, 15, pp. 3-67.

Martin, J.R (1997) Analysis in Genre: Functional Parameters In: Christie, F. and Martin, J. R. (eds) *Genre and Institutions: Social Processes in the Workplace and School*. London: Cassell. pp. 3-39.

Martin, J.R. (2003) Beyond Exchange: Appraisal Systems in English. In Hunston, S. and Thompson, G. (eds.) *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press. pp. 142-175.

Martin, J.R. (2002) Writing History: Constructing Time and Value in Discourse In Schleppergrell, M. J. Colombi, N, C. (eds.) *Developing advanced literacy in first and second languages: Meaning and Power*. New Jersey: Lawrence Erlbaum. pp. 87-118.

Martin, J.R. and D. Rose (2003) *Working with discourse: meaning beyond the clause*. London: Continuum.

Martin, J. R., and White, P. R. (2005) *The Language of Evaluation: Appraisal in English*. Palgrave Macmillan: Hampshire.

Maslow, A. (1970) *Motivation and Personality*. New York: Harper and Row.

Matoesian, G. M. and Coldren Jr., J. R. C. (2002). Language and bodily conduct in focus groups evaluations of legal policy. *Discourse and Society*, 13 (4), pp. 469–493.

Matlock, T. (1989) Metaphor and the Grammaticalization of Evidentials. In: Mushin, I. (ed.) *Evidential and Epistemological Stance: Narrative Retelling*. Philadelphia: John Benjamins. pp. 215-225.

McNeill, D. (2000) *Language and Gesture*. Cambridge: Cambridge University Press.

McNeill, D. (2005) *Gesture and Thought*. Cambridge: Cambridge University Press.

Mehrabian, A., and Ferris, S.R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology*, 31, pp. 248-252.

Mehrabian, A., and Wiener, M. (1967). Decoding of inconsistent communications. *Journal of Personality and Social Psychology*, 6, pp. 109-114.

Mehrabian, A., (1969) *Tactics in Social Influence*. Prentice Hall: New Jersey.

Mehrabian, A. (1971). Verbal and nonverbal interaction of strangers in a waiting situation. *Journal of Experimental Research in Personality*, 5, pp. 127-138.

Mishler, E.G. (1995). Models of Narrative Analysis: A Typology. *Journal of Narrative and Life History*, 5 (2), pp. 87-123.

Miskovic-Lukovic, M. (2009) Is there a chance that I might kinda sort of take you out to dinner? The role of the pragmatic particles kind of and sort of in utterance interpretation. *Journal of Pragmatics*. 41, pp. 602-625.

Mitchell, D. J, Kahn, B.E, and Knasko, S.C. (1995) There's Something in the Air: Effects of Congruent or Incongruent Ambient Odor on Consumer Decision Making. *Journal of Consumer Decision Research*, 22 (9), 229-238.

Morgan, D. (1998) *Planning Focus Groups*. Sage: London

Mozziconacci, S.J.L. (2001). 'Emotion and attitude conveyed in speech by means of prosody'. *For the 2nd Workshop on Attitude, Personality and Emotions in User-Adapted Interaction*, Sonthofen, Germany, July 2001.

Müller. F., E., (1996) *Affiliating and disaffiliating with continuers: prosodic aspects of reciprocity*. Cambridge: CUP.

Müller, C (2004) The palm-up-open-hand. A case of a gesture family? In: C. Müller and R. Posner (eds.), *The semantics and pragmatics of everyday gestures*, Berlin: Weidler Verlag. pp. 233-256

Müller, C., and Posner, R. (2004) (eds.), *The semantics and pragmatics of everyday gestures*, Berlin: Weidler Verlag.

Mushin, I. (2001) *Evidential and Epistemological Stance: Narrative Retelling*. Philadelphia: John Benjamins.

Myers, G.A. (1989) The pragmatics of politeness in scientific articles. *Applied Linguistics*, 10, pp. 1-35.

Myers, G.A. (1996) Strategic vagueness in academic writing. In: Ventola, E. and Mauranen, A, (Eds.), *Academic writing: Intertextual and textual issues* Amsterdam: John Benjamins. pp 3-18.

Myers, G. (1998). Displaying opinions: Topics and disagreement in focus groups. *Language in Society*, 27 (1), pp. 85–111.

Myers, G. (2004) *Matters of Opinion: Talking about Public Issues*. Cambridge: Cambridge University Press.

Myers, G. (2007) Enabling talk: How the facilitator shapes a focus group. *Text and Talk*, 27 (1), pp. 79-105.

Narver, J.C. and Slater, S.F. (1990) The effect of market orientation on business profitability. *Journal of Marketing*, 54, pp. 20–35.

Nassaji, H. and Wells, G. (2000) What's the use of 'Triadic dialogue'? An investigation of teacher–student interaction. *Applied Linguistics*, 21, pp. 376–406.

Nicosia, F. (1966) *Consumer Decision Processes*, Englewood Cliffs: Prentice Hall Publishing.

Norman, D. (2004) *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic Books.

Normann, R. (1984) *Service Management: Strategy and Leadership in Service Businesses*. New York: Wiley.

O' Donnell, M. (2007) CorpusTool Software v 1.14 (downloaded from <http://www.wagsoft.com/CorpusTool/download.html>).

Oliver, D. G., Serovich, J.M. and Mason, T. L. (2005) Constraints and Opportunities with Interview Transcription: Towards Reflection in Qualitative Research. *Social Forces*, 84 (2), pp. 1273-1289.

Olshtain, E. (1989) *Apologies across Languages*. In Blum-Kulka, S. et al. (eds.). *Cross-Cultural Pragmatics: Requests and Apologies*. Norwood, NJ: Ablex, 1989, pp. 155-173.

Oppenheim, A. (1992) *Questionnaire, Interviewing Design and Attitude Measurement* London: Pinter.

Pagano, A. (1994) Negatives in written text. In: Malcolm Coulthard (ed), *Advances in written text analysis*, London: Routledge. pp. 250-265.

Page, R. E. (2003) An analysis of APPRAISAL in childbirth narratives with special consideration of gender and storytelling style. *Text*, 23(2) pp. 211–237.

Painter, C. (2003) Developing attitude: An ontogenetic perspective on APPRAISAL. *Text*, 23 (2), pp. 183–209.

Palmer, F. R. (1986) *Mood and Modality*. Cambridge: Cambridge University Press.

Patton, M. (2002) *Qualitative Research and Evaluation Methods* London: Sage.

Pham, M. T. (1998) Representativeness, Relevance, and the Use of Feelings in Decision Making. *Journal of Consumer Research*. 25, pp. 144-159.

Pham, M. T. (2004). The Logic of Feeling. *Journal of Consumer Psychology*, 14, 360-369.

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/ dispreferred turn shapes. In: J. M. Atkinson and J. Heritage (eds.) *Structures of*

Social Action: Studies in Conversation Analysis,. Cambridge: Cambridge University Press. pp. 57–101.

Precht, K. (2003a) Great vs. Lovely: Stance Differences in American and British English, *Language and Computers* 46, pp. 133-152.

Precht, K. (2003b) Stance moods in spoken English: Evidentiality and affect in British and American conversation. *Text*, 23 (2), pp. 239-257.

Procter, M. (2001) Measuring Attitudes. In: Gilbert, N (ed.) (2001) *Researching Social Life*. London: Sage. pp. 206-243.

Protopapas, A. and Lieberman, P. (1997) Fundamental frequency of phonation and perceived emotional stress. *Acoustical Society of America*, 10 (4), pp. 2267-2277.

Puchta, C. and Potter, J. (2003). *Focus group practice*. London: Sage.

Puchta, C., Potter, J., and Wolff, S. (2004). Repeat receipts: A device for generating visible data in market research focus groups. *Qualitative Research*, 4 (3), pp. 285–309.

Quirk, R. Greenbaum, S., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London: Longman.

Rajagopal, P. and Burnkrant, R. E. (2009) Consumer Evaluations of Hybrid Products. *Journal of Consumer Research*. 36, (2), pp. 232–241.

Rees-Miller, J. (2000) Power, Severity and context in Disagreement. *Journal of Pragmatics*. 32 (8), pp. 1087-1111.

Reynolds, T. J. and Gutman, J. (1988) Laddering Theory: Method, Analysis and Interpretation. *Journal of Advertising Research*, 28, pp. 11-31.

Richins, M. L. (1997) Measuring Emotions in the Consumption Experience. *Journal of Consumer Research*, 24 (9), pp. 127-147.

Sacks, H. (1972) An Initial Investigation of the Usability of Conversational Data for Doing Sociology. In: Sudnow, D. (ed.) *Studies in Social Interaction*, Free Press: New York, pp. 31-74.

Sacks, H. (1974) An Analysis of the Course of a Joke's telling in Conversation. Bauman, R. and Sherzer, J.F. (eds.) *Explorations in the Ethnography of Speaking*. Cambridge: Cambridge University Press. pp. 337-353.

Sacks, H. (1984) On doing 'being ordinary'. In: Atkinson, J and Heritage, J. (eds.) *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press. pp. 413-430.

Sacks, H., Schegloff, E. A. and Jefferson, G. (1974) A Simplest Systematics for the Organisation of Turn-Taking for Conversation, *Language*, 50, pp. 696-735.

Salager-Meyer, F. (1994) Hedges and textual communicative function in Medical English written discourse. *English for Specific Purposes*. 13 (2), pp. 149-170.

Schegloff, E. (1968) *Sequencing in Conversational Openings*. New York: AAA.

Schegloff, E. and Sacks, H. (1973) Opening up Closings. *Semiotica*, 7, pp. 289-327.

Schiffrin, D. (1994) *Approaches to Discourse*. London: Blackwell.

Schuman, H. and Presser, S. (1981) *Questions and Answers in Attitude Surveys*. London: Academic Press.

Searle, J. (1969) *Speech Acts*. Cambridge: Cambridge University Press.

Semenik, R. J. (2002) *Promotion and Integrated Marketing Communications*. Cincinnati, OH: Southwestern Thomson Learning.

Shriberg, E. (1998) Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech*, 41 (3), pp. 439-487.

Sidnell, J. (2010) *Conversation Analysis: An Introduction*. London: Wiley – Blackwell.

Siegel, S. and Castellan, N. J. Jr. (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGrawHill.

Sieglwart, H. and Scherer, K.R. (1995) Acoustic concomitants of emotional expression in operatic singing: the case of Lucia in *Ardi gli incense*. *Journal of Voice*, 9 (3), pp. 249-260.

Silverman, D. (2004) *Qualitative Research; Theory, Method and Practice*. London: Sage.

Silverman, D. (2006) *Interpreting Qualitative Data*. London: Sage.

Simon-Vandenberg, A. (2000) The functions of *I think* in political discourse. *International Journal of Applied linguistics*, 10 (1), pp. 41-63.

Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. (1992) *Advances in Spoken Discourse Analysis*. London: Routledge.

Sinclair, J. (1994). Trust the Text. In: M. Coulthard (Ed.) *Advances in Written Text Analysis* New York: Routledge. pp. 12–25.

Sinclair, J. (1996). The Search for Units of Meaning. *Textus*, IX, pp. 75–106.

Sinclair, J. M. and Coulthard, M. (1975) *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. Oxford: Oxford University Press.

Siraj-Blatchford, I. (1995) Kelly's repertory grid: a technique for developing evaluation in design and technology. (Unpublished Paper Available at <https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/1526>).

Sivacek, J. and Crano, W.D. (1982) Vested interest as a moderator of attitude-behaviour consistency. *Journal of Personality and Social Psychology*, 43, pp. 210-21.

Smith, D.V.L and Fletcher, J. H. (2006) *The Art and Science of Interpreting Market Research Evidence*. Chichester: Wiley.

Sornig, K. (1977) Disagreement and contradiction as communicative acts. *Journal of Pragmatics*, 1(4), pp. 347-373.

Sorrentino, R.M. and Higgins, E.T. (Eds) (1986) *The Handbook of Motivation and Cognition*. Chichester: Wiley.

Spencer-Oatey, H. (1996) Reconsidering Power and Distance. *Journal of Pragmatics*, 26 (1), pp. 1-24.

Stewart, D. W., Shamdasani, P. N., and Rook, D. (2007) *Focus Groups: Theory and Practice (Applied Social Research Methods)*. London: Sage.

Stubbs, M. (1995) Collocations and semantic profiles: on the cause of the trouble with quantitative methods'. *Functions of Language*, 2 (1), pp. 23-55.

Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Blackwell, Oxford.

Stubbs, M. (2006) *Corpus analysis: The state of the art and three types of unanswered questions*. London: Equinox.

Summers, G. (ed.) (1970) *Attitude Measurement*. Chicago: Rand McNally.

Swientek, B. (2001) Uncanny Developments, Beverage Industry. *Journal of Marketing Theory and Practice*, 10 (4), pp. 58-68.

Tapper, J. (1996) Exchange patterns in the oral discourse of international students in university classroom. *Discourse Processes*, 22, pp. 22-55.

Taylor, J. R. (1995) *Linguistic Categorisation*. Oxford: Oxford University Press.

Thompson, G. (1996) *Introducing Functional Grammar*. London: Arnold.

Thornborrow, J. (2002) *Power Talk: Language and Interaction in Institutional Discourse* London: Prentice Hall

Tottie, G. (1987) Rejections, denials and explanatory statements - a reply to Fretheim. *Studia Linguistica*, 4 (1), pp. 154-163.

Tottie, G., and Hoffman, S. (2006) Tag questions in British and American English. *Journal of English Linguistics*, 34 (4), pp. 283-311.

Turley, L. W. and Milliman, R. E. (2000) Atmospheric Effects on Shopping Behaviour: A Review of the Experimental Evidence. *Journal of Business Research*, 49, pp. 193-211.

Underwood, R.L. and Klein, N.M. (2002) Packaging as brand communication: effects of product pictures on consumer responses to the package and brand. *Journal of Marketing Theory and Practice*, 10 (4), pp. 58-68.

- Valentine, V and Gordon, W. (2000) The 21st century consumer - A new model of thinking, *Proceedings of the Market Research Society Conference*.
- Van der Auwera, J. and Plungian, V. (1998) Modality's semantic map. *Linguistic Typology*, 2 (1), pp. 79-124.
- Van Dijk, T. (1999) Critical Discourse Analysis and Conversation Analysis. *Discourse and Society*, 10, pp. 459-460.
- Van Kleef, E., Van Trijp, H. and Luning, P. (2005) Consumer research in the early stages of new product development: a critical review of methods and techniques. *Food Quality and Preference*, 16 pp. 181–201.
- Verschueren, J. (1999) *Understanding Pragmatics*. London: Arnold.
- Volosinov, V. N. (1995). *Marixism and the Philosophy of Language*. Translated by L Matjka and I. R. Titunik. London: Routledge.
- Von Hippel, E. (1988) *The sources of Innovation* Oxford: OUP.
- Watt, R. J. (2003) *Politeness*. Cambridge: Cambridge University Press.
- White, P. R. R. (1998) Telling media tales: The news story as rhetoric. (unpublished PhD dissertation. University of Sydney, Sydney) Available from www.grammatics.com.
- White, P. R. R. (2003) Beyond modality and hedging: A dialogic view of the language of intersubjective stance. *Text*, 23 (2), pp. 259-284.
- Wicker, A. W. (1969) Attitudes Versus Actions: The relationship of verbal and overt behavioural responses to attitude objects. *Journal of Social Issues*. 25, pp. 41-78.

Wicker, A. W. (1971) An Examination of the “other variable” explanation of attitude-behaviour inconsistency. *Journal of Personality and Social Psychology*. 19, pp. 18-30.

Willet, T. (1988) A cross linguistic survey of the grammaticalisation of evidentiality’. *Studies in Language*, 12 (1), pp. 51-97.

Willis, D. 1992. ‘Caught in the act: using the rank scale to address problems of delicacy’. In: Malcolm Coulthard (ed), *Advances in written text analysis*, London: Routledge. pp. 111-122.

Whitsitt, S, (2005) A Critique of the Concept of Semantic Prosody. *International Journal of Corpus Linguistics*, 10, (3), pp. 283-305.

Wodak, R. and Meyer, M. (2001) *Methods of Critical Discourse Analysis*. London: Sage

Wolfson, N. (1976) Speech events and natural speech: some implications for sociolinguistic methodology. *Language in Society*, 5, pp. 189-209.

www.grammatics.com (2005).

Yngve, V. (1970). On getting a word in edgewise. In: *Papers from the 6th Regional Meeting, Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.

Yngve, V. and Wasik, Z. (2006) *Hard Science – Linguistics*. London: Continuum.

Appendices

Discussion Guide for Initial Packaging Product Group: Stage 1

Things to remember:

Make sure the video is playing!!!

Make sure they all sign the sheet to say they've been paid

Make sure they've signed the consent form

Focus groups comprised of 6 18-30 year old, male, Lynx / body spray users.

The aim is to explore general perceptions of different deodorant brands and what role the packaging plays in creating and maintaining those perceptions

Standard focus group set up with different stimuli being used to help elicit opinions.

Stimuli to be used:

- Concept boards
- Current packaging designs from differing markets

Ok thanks for coming, I'm not sure what you've been told but I presume that you're aware that we're going to be talking about deodorants and body sprays and packaging and brands. I'm carrying out some research independently of Unilever so feel free to make negative comments about Unilever products, don't feel like you have to be polite about them. I'm after your honest opinions about products and ideas about packaging.

We're quite a big group so while I want everyone to get involved, it would be great if you could try not to talk over each other too much as it's important that I hear everyone's opinions.

I should just mention that this is being recorded and there might be people watching behind the mirror, though they're watching me rather than anything you might be saying.

Ok so just to get us started can just take a minute or two to introduce yourself to the person sitting next to you and then you can introduce each other to the group- so can you just say obviously who you are, what you do and what deodorants or sprays you're using at the moment.

Why do you use your current deodorant?

- Can you describe what it looks like to me- do you (dis)like it?
- Do you only use it? How long have you used it for? What did you use before?
- Would you consider changing? What could make you change?
- Which varieties / fragrances do you use? Why?
- Do you use more than one? If so, why? Where?

BRING OUT THE PRODUCTS STIMULI

How much impact do you think the design of a product has on whether or not you buy it?

- Do you think that the design / packaging affects how much you'd be prepared to pay for a product?
- Do the design / packaging affect how well you think something is going to work?

Can you think of a product from any market that you particularly like the packaging?

Can you think of a product that you've bought because you liked the packaging rather than because you liked the product?

Can you think of a product that you haven't bought because you really disliked its look, even if it worked really well?

What kind of product is this most likely to be a consideration for?

Which features do you like and dislike from the products on display?

- Why?
- What do they tell you about the product?
- Why do you like that?

Product perceptions to explore

- What kind of person would you expect to see using the product?
- How well would you expect the product to work?
- Is it an expensive or cheap product?

What about the product causes these perceptions?

- Shape?
- Colour?
- Texture?
- Size?
- Combinations of them?

1. Are there any products from other markets, which have designs that you like?
 - a. This could be anything from other personal care goods, food containers, cleaning products- any packaging that you like the look of.
2. Are there any products from other markets, which have triggers / outputs that you like?
 - a. Again this could be from any product that involves some kind of action to get a product out of a pack - air fresheners, olive oils, after shaves, perfumes etc.

3. Do you think that these could be used for deodorants? Would they be appealing?
 - a. If not, why not- what would stop it from working?

4. Do you see the pack as being made up of parts, or as just a whole?
 - a. How many different parts to the pack are there?

5. If you had to divide it up into different parts how would you divide it up

BRING OUT THE CONCEPT BOARDS

What do you think of these as ideas?

Which do you like most, least?

What do you like, or dislike about them?

Now what I want you to do is to think about all that we've talked about today and about the different designs that there are and I want you to just sketch a design that you think would look good for a deodorant and that you would like to see made.

Not sure if you were made aware of this but some of you may be asked to come back in a week or two to take part in a session where we will explore these ideas further and actually attempt to create the design that you have just drawn.

Thank you

II Discussion Guide for CAD Groups: Stage 2

An individual session with CAD operator and moderator to explore different design possibilities that appeal to the consumer.

Six of the participants from the initial focus groups will be chosen to come back and take part in these sessions.

Their drawings from the groups will be used as the starting point and they will then work with the CAD operator to play around with this design to create their ideal packaging solution.

Initial talk through of the aims of the session, i.e. Getting a CAD design that represents their ideal deodorant pack for their brand.

Start off by asking them to describe mystery object in a covered box to get them used to using that kind of descriptive 'shape' language. See how close their description, as modelled by the CAD designer, is to the actual object.

Then bring up the first draft of their design on the CAD software

Start off by asking them to describe the pack:

- What it looks like:
 - Why they chose that design and what they like about it.
 - How is it different to their current pack and why it is better

- What it reminds them of?
 - Packs from other genres?
 - Where did the idea come from?

- Physical aspects of it?
 - How big is it? Why? What happens if we make it smaller?
 - How long is it? Why? What happens if we make it shorter
 - What should the surface feel like? Why? What happens if we change that?

- What are the features of this pack?

- How is it an improvement on the current design?

Ask them to describe what they think the Lynx brand is and then if this is for Lynx (their preferred brand) what about the design relates to the brand? How do the features tie in with that?

What about it would appeal to Lynx users?

Then as they're talking through changes, ask them how that changes their perceptions and why they wouldn't change it in the other direction.

How do the changes relate to the brand?

You've done it as Lynx, if it was Tesco's own brand, would it look different? How would it look different? How would that make you feel about it?

You said it was X what about it makes it look X?

If it was taller would it still look X?
 fatter
 wider
 shorter
 curvier
 straighter-lined

General Questions to be asked throughout the session:

6. Do you think that this pack could be used for both male and female fragrances? If not why not?
7. Could this pack be used for any other products apart from deodorant? If so, why and what? If not, why not?

8. How do you think this pack would be to use? Robust, Easy?
9. Would you see this pack as being for a young or old person?
10. Would you see this pack as being for a cheap or premium product?

Other questions?

11. Are there any products from other markets, so not just deodorants or cosmetics, that have designs that you like?
12. Are there any products from other markets, which have designs that you like?
13. Are there any products from other markets, so not just deodorants or cosmetics, that have triggers / outputs that you like?
14. Are there any products from other markets, which have triggers / outputs that you like?
15. Do you think that these could be used for deodorants? Would they be appealing?
16. Do you see the pack as being made up of parts, or as just a whole?
17. If you had to divide it up into different parts...
18. How would you divide your can into different parts...

19. How do the different parts influence your overall perception

20. How would you describe the different parts?

III Discussion Guide for Paired Depths Interviews: Stage 4

These are paired depth interviews with the aim of exploring the participant's opinions with regards to the shape and design of the prototypes. The stimuli will be the prototypes that have been made as a result of the previous three stages of the Packaging Project.

Ensure that the participants are aware that the session is being recorded and that there may be people watching behind the screen. Reassure them that the recordings are for just to aid my memory and that it is my moderating skills that are being watched rather than them. Check that they are happy to continue.

Ask them to introduce themselves to each other and then to me. Introduce self.

Explain that we're looking at prototype designs for deodorant bottles and that they are all unbranded and without any moving parts. Explain that it is the actual shape and design of the packs that we are looking at.

There are two stages to what we're going to be doing and for the first one I'm going to show you one design at a time and I want you to describe it to me

There aren't any right or wrong answers so don't worry if you think what you are going to say is wrong in anyway

- Does it make you feel anything?
- Does it remind you of anything?
- What do you like about it?
- What do you dislike about it?

- Do you see it as a whole or as parts?
- If parts what are they?
- What kind of product do you think it is for?
- What kind of person would use this?
- Who do you think that this is aimed at?
- Do you think this would suit one brand more than another?
- What would you change about it?

If you disagree with what the other person has said then please make sure you comment- constantly check to see if they agree with each other if not explore the differences in their opinion. See if they understand the other view point.

Lynx Perspective Questions

We talked before that this might be more suitable for one brand than another- what I want do now is look at that in more detail.

F – is it fair to presume that neither of you use Lynx?

Are you aware of

Triadic Sorting

I want you to choose two designs that have something in common that you like and then choose another one that doesn't have that feature.

Laddering Technique

- What is the shared feature that you like?
- What is it that you like about that?
- Why is that important to you?
- Why don't you like the fact that it is missing from the other design?

Bring out Current Packs

Do you think any of these would be more suited to any of these brands or products?

IV Instructions for Inter-raters

Background Information: My work

The aim of my research is to show that linguistics tools (In this case Appraisal Theory) can be used to provide insights and add clarity to the interpretation of evaluative language. The context for my research is market research focus groups, with a subject matter of new product development and the evaluation of products and prototypes.

Appraisal Theory

Appraisal Theory is a system of analysis that has been developed over the last few decades and can be seen as an extension of Halliday's Systemic Functional Grammar. Appraisal Theory makes use of three sub-systems that can be applied to texts to analyse them in terms of different areas of interest. The three areas are; Attitude, Engagement and Graduation.

Attitude

The questions in brackets should be used as a test to check that the right tag has been applied to the transcript.

	Positive	Negative
Reaction – Impact	Arresting, captivating,	Dull, boring, tedious,

<i>(Did it grab me?)</i>	engaging, fascinating	uninviting
Reaction – Quality <i>(Did I Like it?)</i>	Lovely, beautiful, appealing, splendid	Plain, ugly, repulsive, revolting
Composition – Balance <i>(Did it fit together?)</i>	Balance, harmonious symmetrical	Unbalanced, discordant, distorted, stretched
Composition – Complexity <i>(Was it hard to follow)</i>	Simple, elegant, intricate, detailed	Simplistic, monolithic, extravagant
Valuation <i>(Was it worthwhile?)</i>	Original, unique, challenging	Insignificant, shallow, conservative

This can also be extended to multiword phrases. ‘That would stand out on the shelf’ is tagged as [ATTITUDE: APPRECIATION: REACTION: IMPACT] as it clearly answers the question ‘*Did It grab me*’. Similarly, comments regarding the physical nature of the pack are tagged as [ATTITUDE: APPRECIATION: COMPOSITION: COMPLEXITY]. ‘*It’s too big*’, ‘*it’s very bulky isn’t it*’, ‘*that’s massive*’.

Although a comment such as ‘*I really like that*’ seems to answer the question ‘*did I like it?*’ it is not just tagged as [ATTITUDE: APPRECIATION: REACTION: QUALITY]. Instead it is also tagged as [AFFECT: POSITIVE] as it is an emotional disposition.

The polarity of the tagging will be obvious in many cases. However in some cases it might not be so clear so it is necessary to look to the surrounding evaluations to gain an indication of the speaker’s overall evaluation of the object. Previous, or future, comments on the same aspect of the pack can also give a greater level of indication as to the intended polarity.

Engagement

These are any meanings by which speakers either acknowledge or ignore the diversity of view-points put at risk by their utterances and negotiate an interpersonal space for their own positions within that diversity. Shows the level to which they adhere

to or disagree with the expected 'shared community' evaluation / appraisal of an object or event. In other words, are they saying that their opinion is definitely right or are they accepting that some people might disagree? This is the difference between a Monoglossic evaluation and a Heteroglossic evaluation.

For example:

- modals of probability - *perhaps, it may..., I think..., surely*
- reality phase - *it seems,*
- attribution (hearsay/projection) - *his alleged ..., informed sources report ..., scientists have found evidence suggesting that,*
- proclamation - *In fact, I am compelled to conclude ..., It is true, we do have a small black and white cat*
- expectation - *predictably, of course,*
- counter-expectation - *amazingly*

I've simplified how I use Engagement and only measure whether the speaker has increased or decreased their Engagement with the evaluation. Increase would include things such as 'definitely' 'obviously', 'certainly' Decrease would be things like *perhaps, possibly, maybe, it looks/seems/feels, I think, I don't know* etc

Graduation

Values by which (1) speakers graduate (raise or lower) the interpersonal impact, force or volume of their utterances, and (2) by which they graduate (blur or sharpen) the focus of their semantic categorisations.

1. (FORCE) *slightly, somewhat, very, completely*
2. (FOCUS) *I was feeling kind'o woozy, they effectively signed his death warrant; a true friend, pure folly*

If the force is strengthened through words like 'that's *really* bad' this would be Graduation: Force: Heighten Intensity. If the force is weakened with something like 'that's *quite* bad' this would be Graduation: Force: Lower Intensity.

Rules for determining which tag is applied

Attitude

- Is it related to their visual response to the pack?
 - Is it related to its visual impact? [ATTITUDE: APPRECIATION: REACTION: IMPACT]
 - Is it positive or negative? [NEGATIVE/ POSITIVE]
 - Is it related to its quality? [ATTITUDE: APPRECIATION: REACTION: QUALITY]
 - Is it positive or negative? [NEGATIVE/ POSITIVE]
- Is it related to the composition to the pack?
 - Is it related to how it looks? [ATTITUDE: APPRECIATION: COMPOSITION: BALANCE]
 - Is it positive or negative? [NEGATIVE/ POSITIVE]
 - Is it related to how easy it looks to follow/use? [ATTITUDE: APPRECIATION: COMPOSITION: COMPLEXITY]
 - Is it positive or negative? [NEGATIVE/ POSITIVE]
- Is it their overall valuation? [ATTITUDE: APPRECIATION: VALUATION]
 - Is it positive or negative? [NEGATIVE/ POSITIVE]

Graduation

Force

Does it mitigate or hedge the position? [GRADUATION: FORCE←]

Does it strengthen the evaluation? [GRADUATION: FORCE←]

Focus

Does it blur the focus of the semantic category or group? [GRADUATION: FOCUS←]

Does it sharpen the focus of the semantic category or group? [GRADUATION: FOCUS→]

Engagement

Does it seem as if the speaker is admitting that others might think differently?

[ENGAGEMENT: DECREASE←]

Does it seem as if the speaker is adamant that their view is the only one that could be right? [ENGAGEMENT: INCREASE→]

V Transcripts

The transcripts are on the accompanying CD as part of the CorpusTool folder